Spatial Content Understanding of Very High Resolution Synthetic Aperture Radar Images

DISSERTATION
zur Erlangung des Grades eines Doktors
der Ingenieurwissenschaften

vorgelegt von
Jagmal Singh, M.Sc.

eingereicht bei der Ingenieur- und Naturwissenschaften Fakultät
der Universität Siegen
Siegen 2014

Stand: Juni 2014

Promotionskommission:

1. Gutachter: Prof. Dr. Otmar Loffeld
2. Gutachter: Prof. Dr. Mihai Datcu

1. Mitglied der Promotionskommision: Prof. Dr. Horst Bessai
2. Mitglied der Promotionskommision: Prof. Dr. Uwe Stilla

Vorsitz des Prüfungskommission : Prof. Dr. Horst Bessai

Tag der mündlichen Prüfung: 30. Mai 2014

Zusammenfassung

Die Verfügbarkeit von großen Mengen von Fernerkundungsbildern mit einer Auflösung im Meterbereich, wie sie die neueste Generation von Radar-Satelliten mit synthetischer Apertur (SAR) erzeugt, bedingt auch neue Untersuchungen. Um zutreffende Bilder in großen Datenbanken zu suchen und daraus abzurufen, werden neue Verfahren zur automatischen Analyse, Interpretation und Indexierung von SAR-Bildern benötigt. Die bisher dazu entwickelten Methoden basieren auf dem Verstehen der Speckle-Eigenschaften in SAR-Bildern. Der Schwerpunkt liegt dabei auf der modellbasierten Schätzung von Texturparametern im Amplitudenraum von SAR-Bildern, wie zum Beispiel bei parametrischen Gibbs-basierten Methoden in einem Bayes-Modell. Diese Methoden haben sich als sehr erfolgreich erwiesen für SAR-Bilder mit einer Auflösung im Bereich von $10 - 50$ Metern unter der Annahme von stationären Signalen in einem Suchfenster mit passender Größe.

Die Herausforderung bei SAR-Bildern mit einer Auflösung im Meterbereich liegt im Vorhandensein von sehr vielen isoliert instationären Details, wo eine modellbasierte Parameterschätzung ungenaue Ergebnisse liefert. Diese Einschränkung führt uns zur Hinwendung auf nicht-parametrische Strategien mit Nutzung der Phaseninformation zur Transformation von SAR-Bildern in einen geeigneten Raum. Ein grundlegender Beitrag dieser Arbeit ist die Darstellung der Vorteile und der Bedeutung der Phaseninformation in komplex-wertigen SAR-Bildern gegenüber der Verwendung des reinen Amplitudenraums für solche Strategien.

Wir untermauern die Bedeutung der Phaseninformation mit einem Vorschlag für eine Methode zur Zerlegung von Bildern in mehrere Sub-Looks (MSLD). Diese Methode erzeugt Hyperbilder aus der spektralen Analyse von komplex-wertigen SAR-Bildern und erlaubt die visuelle Untersuchung von Objekten. Darauf aufbauend wird gezeigt, dass eine aus Chirplets abgeleitete Transformation die gebrochene Fourier-Transformation (FrFT) - einen echten und für SAR bedeutsamen Multiskalenansatz darstellt, bei der die Skalierung in der Phase durchgeführt wird. Ein vorgeschlagener nicht-parametrischer Deskriptor zur Merkmalsbeschreibung, der auf statistischen Maßen zweiter Ordnung (logarithmischen Kumulanten) beruht und durch die Größe des Amplitudenraums der FrFT-Koeffizienten geschätzt wird, zeigt im Merkmalsraum eine erhöhte Trennbarkeit für eine verbesserte Indexierung.

Zur Validierung des vorgeschlagenen nicht-parametrischen Verfahrens auf Grundlage der FrFT wurde zum Vergleich mit bereits existierenden Methoden eine experimentelle Datenbasis für Leistungstests mit Single Look Complex (SLC) TerraSAR-X-Bildern im Spotlight-Modus generiert. Zur Beurteilung und zum Vergleich der untersuchten Algorithmen schlagen wir eine robuste methodologische Klassifizierungs-Umgebung vor.

Abstract

Availability of large amounts of very high resolution (metric-resolution) remote sensing images from the last generation synthetic aperture radar (SAR) satellites is attracting new studies. In order to search and retrieve relevant images from large-scale databases, new techniques for automatically analyzing, interpreting and indexing SAR images are required. The methods developed for this purposes in the past were based on the understanding speckle characteristics in SAR images. The focus has been generally on the model-based textural parameter estimation in the amplitude-envelope of SAR images, such as parametric Gibbs-based methods in the Bayesian framework. Such methods were largely successful under the assumption of stationarity of the signal in an analyzing window of convenient size on images with resolution of the order of tens of meters.

The challenge we encounter in metric-resolution SAR images is the presence of a very high order of details encapsulating a non-stationarity, where model-based parameter estimation becomes inaccurate. This constraint encourages us to focus on non-parametric strategies while employing phase information to transform SAR images in a suitable space. Demonstrating the advantages and relevance of the phase information embedded in complex-valued SAR images over the use of the mere amplitude-envelope for such strategies is an underlying contribution of this thesis.

The importance of phase information is advocated with a proposed method of multiple sublook decomposition (MSLD). This method generates hyper-images from the spectral analysis of complex-valued SAR images enabling the visual exploration of targets. Subsequently, a chirplet-derived transform- the fractional Fourier transform (FrFT) has been found to be a true SAR relevant multi-scale approach, where scaling is carried out in the phase. A proposed non-parametric feature descriptor based on the use of second-kind statistical measures (logarithmic-cumulants) estimated over the amplitude-envelope of the FrFT coefficients exhibits enhanced feature space separability for improved indexing.

An experimental benchmarking database is generated on single look complex (SLC) spotlight mode TerraSAR-X images for the validation of the proposed FrFT-based nonparametric technique in comparison to the existing methods. A robust methodological classification framework has been proposed for the evaluation and comparison of the studied algorithms.

Acknowledgements

First of all, I would like to express my gratitude towards Prof. Mihai Datcu, my supervisor at *German Aerospace Center* (DLR), for his continuous scientific guidance, support, and motivation. My research work in the present form could not have been possible without his scientific enthusiasm and rigor. I am also grateful to Prof. Otmar Loffeld from *Universität Siegen* for giving me an opportunity to carry out this research work under the *International Postgraduate Program Multi Sensorics* and carefully reviewing my thesis with valuable suggestions.

I would like to thank Prof. Peter Reinartz, head of department of *Photogrammetry and Image Analysis* (PBA), and Prof. Richard Bamler, director of *Remote Sensing Technology Institute* (IMF), for providing me a very stimulating working environment along with motivation and suggestions over all these years. I would also like to thank Dr. Holger Nies from *Universität Siegen* for his support in all administrative matters.

Thanks to *German Academic Exchange Service* (DAAD) for financially supporting my research work in the frame of DLR-DAAD Doctoral Fellowship.

I would like to thank Daniela, Daniele, Gottfried, Dusan, Anca, Shiyong, Carmen, Corina and Octavian for valuable technical discussions we had in the meeting room during the day, or at beer gardens of Munich during the evenings. Thanks to Jakub, Janja, Esteban, Octavio, Kanika and Olena for never leaving me alone for lunch at work and also for the much needed cups of coffee we had together before going back to research in office. I would also like to thank all my colleagues at IMF-PBA for daily exchange of ideas and facilitating a very congenial work environment. Many thanks to Gopika, Vaidehi and Daniele for carefully proofreading this report. My special thanks go to Sunil, Sahil, Abhijit, Dorota, Jan, Sebastian Gaigle, Stefan and Sebastian Türmer, who never allowed me to forget that there is also a world beyond the doctoral research.

Finally, my deepest gratitude to Susanne, and my whole family in India, who supported me all possible ways to keep me motivated during these years of research.

Dedicated to my loving parents

Contents

List of Figures

Chapter 1

Introduction

One of the goals of signal processing is to extract information embedded in the signals by various possible means. Choosing tractable techniques or developing new methods for analysis and for information extraction from real world signals are not easy tasks. In a similar way, the information extraction from images, which are two dimensional signals, is dependent upon various characteristics such as sensor and resolution. For the case of remote sensing images, the automated methods of information retrieval have always been a challenge. Tremendous amounts of remote sensing images with ever increasing resolution from recent and upcoming synthetic aperture radar (SAR) satellite has opened new challenges and new application areas. In order to search and retrieve relevant images from large-scale databases, we require new techniques for automatically analyzing, interpreting and indexing images, to be used in content based image retrieval (CBIR) systems for SAR images. CBIR systems based on the spatial content of images are already widely applicable in optical remote sensing images (Shyu *et al.* [2007]), as well as to photography pictures (Smeulders *et al.* [2000]).

The problem of spatial content understanding in SAR is dealt differently than passive optical remote sensing images and photography pictures, as SAR is not just an imaging sensor but actually an active coherent device. SAR works on the principle of transmitting coherent electromagnetic pulses and recording the amplitude as well as the phase information of the backscattered signal. One peculiar characteristic of the SAR images is that they are corrupted by multiplicative granular noise, called *speckle* (Touzi [2002]). The methods developed in the past for SAR image processing have started from understanding the speckle characteristics. Development of de-speckling algorithms lead to information extraction through model-based textural parameter estimation, such as the parametric Gibbs-based method in the Bayesian framework. These methods typically focused on the textural parameter estimation for amplitude-envelope of SAR images.

Model-based textural parameter estimation

In the model-based texture parameter estimation in the Bayesian framework, the assumption of the *likelihood* probability density function (pdf) is the speckle noise, and the *prior* pdf is the underlying response from the scatterers. The *likelihood* pdf is generally characterized by a Gamma distribution, while the choice of *prior* is a critical issue. The estimated model-parameters can be used as a descriptor of the image, under the assumption that the assumed *prior* pdf is accurate and plausible. In Walessa & Datcu [2000], Gauss-Markov random fields (GMRF) has been proposed as *prior* pdf, whereas in Hebar *et al.* [2009] autobinomial model (ABM) is advocated as a better *prior* pdf for amplitude SAR images. Both methods belong to

the family of Gibbs random fields (GRFs) methods. In this thesis we will do a critical analysis of GMRF and ABM, for spatial content understanding in SAR images.

In general, an image is characterized by its primitive features such as texture (Tuceryan & Jain [1998]), color, and shape (Jain & Vailaya [1996]) to describe its spatial content. This general representation is applicable to multimedia images, as well as optical and SAR satellite images. Gibbs-based parametric methods characterizes the textural information only in the amplitude SAR images, ignoring the phase information. Such methods are successful under the assumption of stationarity of the signal under consideration within a small analyzing window on images with spatial resolution of the order of tens of meters. The challenge in metric-resolution SAR images is their high amount of details encapsulating a non-stationarity, which causes model-based parameter estimation to be inaccurate.

Image representation in a transformed domain

Understanding of SAR images is sometimes not possible using parametric approaches in its spatial representation, as is the case with Gibbs-based methods. Some important features of the SAR signals are more easily characterized in the frequency domain. In order to have more flexibility, the joint use of time and frequency i.e. joint time-frequency analysis (JTFA) is widely preferred. The classical examples of the JTFA may include linear algorithms such as short-time Fourier transform, Gabor expansion or quadratic transforms such as Wigner-Ville decomposition. The use of JTFA methods specific to SAR image analysis has been summarized in Chen *et al.* [2003]. Generally such methods are also applied to amplitude SAR images only, but the use of complex-valued SAR images for spectral analysis has also been discussed (Bertrand & Bertrand [1996], Souyris *et al.* [2003], Colin *et al.* [2004], and Spigai *et al.* [2011], etc.).

Demonstrating the advantage and relevance of the phase information embedded in complex-valued SAR images over the use of the mere amplitude-envelope for such strategies is an underlying contribution of this thesis. The importance of phase information is advocated with a proposed method of **multiple sublook decomposition** (MSLD Singh & Datcu [2012c]). The MSLD method is an extension of SAR relevant sub-look decomposition (Spigai *et al.* [2011]), which generates hyper-images using the spectrum of the complex-valued SAR images in order to provide a visual exploration tool for SAR analysts. Motivated by the simulation results of MSLD, we extend the multi-scale approaches for complex-valued SAR images with the **fractional Fourier transform** (FrFT) (Namias [1980], Ozaktas *et al.* [1996]), the use of which has been up to now limited to moving target detection in SAR images (Sun *et al.* [2002]). The FrFT, which is a chirplet-derived linear transform, is found to be a true SAR relevant multi-scale approach. Here the scaling is carried out in phase and the scale factor is controlled by the transform order of the FrFT. This multi-scale approach characterizes not only the texture, but also the backscattering behavior of scatterers on the ground dependent on the material's dielectric properties, its orientation on ground, and the size of the scatterers compared to the wavelength of the transmitted electromagnetic wave.

In order to generate a compact feature descriptor on the different sub-bands in a multi-scale approach, we then propose the use of **second-kind statistical measures** (logarithmic-cumulants Nicolas [2002]), which provide a simple and efficient alternative to the moments with an improvement in the feature space separability (Singh & Datcu [2013]). The limitation of the traditional multi-scale approaches for the complex-valued SAR images will also be discussed in the thesis.

Pixel-based classification versus patch-oriented image categorization

The spatial content understanding provides a general framework for analyzing, interpreting and indexing SAR images. The applicability of the discovered knowledge is largely dictated by the stationarity of the signal, i.e. the homogeneity of the texture over a considered analyzing window. With relatively homogeneous regions, pixel-based classification is possible with good accuracy, as in low-resolution SAR images. The final result in pixel-based classification/clustering is image pixels classified/clustered into different regions such as water, snow/ice, urban areas or natural areas, with supervised or unsupervised means for large-scale land-cover and crop monitoring, sea-ice studies etc.

With the advent of high-resolution and very-high-resolution SAR images from satellites like TerraSAR-X, TanDEM-X, COSMO-SkyMed, Radarsat-2, etc., interest is now on patch-oriented image categorization (Popescu *et al.* [2012]). For patch-oriented image categorization, our objective is not to classify some regions of the image, but the whole image-patch into an object category. The size of the image patch should be large enough to contain the contextual information needed to define an object's structure used to define a particular category. This application area is gaining importance due to the increase in the details available in a SAR image patch.

These two general application areas, i.e. pixel-based classification and patch-oriented image categorization will be discussed in detail in this thesis.

Evaluation strategy on a reference database

Apart from understanding the different techniques, it is important to have a quantitative analysis of the relevant algorithms for the considered applications. Thus, we also propose a robust **methodological classification framework** (Singh & Datcu [2012a]) under which a qualitative and quantitative analysis of both existing and newly proposed algorithms will be carried out.

For quantitative analysis, an experimental database for pixel-based classification has been generated from single look complex (SLC) spotlight-mode TerraSAR-X images consisting 12 homogeneous and well textured metric-resolution SAR sub-images. The k-nearest neighbors (k-NN) supervised classification has been carried out with 1% and 10% training samples, with an analyzing window (aw) of size 16×16, 32×32, and 64×64 pixels. A second database for patch-oriented image categorization evaluation consist of 2000 image-patches belonging to 20 well-defined objects/texture categories, with 100 patches for each category. The size of each image patch is 200×200 pixels, and each patch covers approximately 200^2 m^2 of ground. The k-NN supervised classification has been carried out with 10% training samples.

Discussion of results with important observations on selected existing methods and newly proposed methods lead us to several important conclusions in spatial content understanding of SAR images.

Part I

Theoretical Background

Chapter 2

SAR Basics

The objective of this chapter is to provide a basic understanding of the signal processing and statistical aspects in synthetic aperture radar (SAR) images, which are essential to form a basis for spatial content understanding in SAR images. The basic imaging principles of a side-looking radar and information about the data acquisition geometry along with the definition of these geometry related terms are presented. In the geometrical perspective, the basic scattering mechanism occurring over the ground depending upon the microscopic and macroscopic properties of the scatterers on the imaged surfaces and the geometrical distortions inherent to SAR images due to the side-looking geometry are discussed.

The topics related to signal processing in SAR i.e. range compression and synthetic aperture by following an approximated rectilinear geometry are presented. The basic overview to achieve an adequately high-resolution in range direction with pulse compression and in azimuth direction by generation of the azimuth chirp in the synthetic aperture is presented. With the discussion of the azimuth chirp, it is pertinent to discuss the concept of azimuth splitting which becomes very prominent in case of high-resolution SAR using broad bandwidth as in TerraSAR-X. This discussion will form the basis for the time-frequency analysis methods proposed in this thesis. The issue of Doppler phase de-ramping for Spotlight mode SAR is also presented.

Later, the statistical modeling of SAR images with the assumption of multiplicative noise-like granular aspect i.e. speckle in SAR images is discussed. This forms the basis of the statistical analysis methods discussed in this thesis. Simple probability density functions (pdfs) for amplitude and intensity SAR images are discussed followed by a short discussion on the more accurate pdf characterization.

We will close this chapter as well as the following chapter in Part-I and Part-II with a short summary as the last section.

2.1 Principles of SAR imaging

History of Radar (acronym for **RA**dio **D**etection **A**nd **R**anging) can be traced back to 1904 when a German inventor *Christian Hülsmeyer* build a simple ship detection device "telemobiloscope", intended to help avoid collisions in fog. Due to technological enhancements in the domain of antenna design and signal processing, it gave birth to synthetic aperture radar (SAR), a revolution in the field of remote sensing for several applications. Introduction of the concept of synthetic aperture, first published by *Carl A. Wiley* in 1953, made it possible to achieve a very-high-resolution in both azimuth and range direction as compared to real aperture radar.

SAR images has many merits over the optical images. An optical sensor, being a passive sensor, has the limitation that it works only at daytime and when the environment is cloud free. SAR has overcome these limitations of passive optical sensors, as it is capable of imaging at any time and any season. Due to this, the use of SAR images has been increased tremendously in several applications. Its sensitivity to target parameters such as dielectric properties and scene characteristics such as large and small scale geometry gives it an edge over optical image data in a number of applications (Curlander & McDonough [1992]).

SAR images are inherently bi-dimensional complex-valued signals and reveal structures both in amplitude and phase. For the ease of interpretation, usually amplitude envelope of the complex-valued image is utilized. Figure 2.1 shows an example of the amplitude, real component, imaginary component and phase information for SAR images and its visual comparison with an optical image of same area. The amplitude-envelope of the complex-valued spectrum of the considered complex-valued SAR image is also presented.

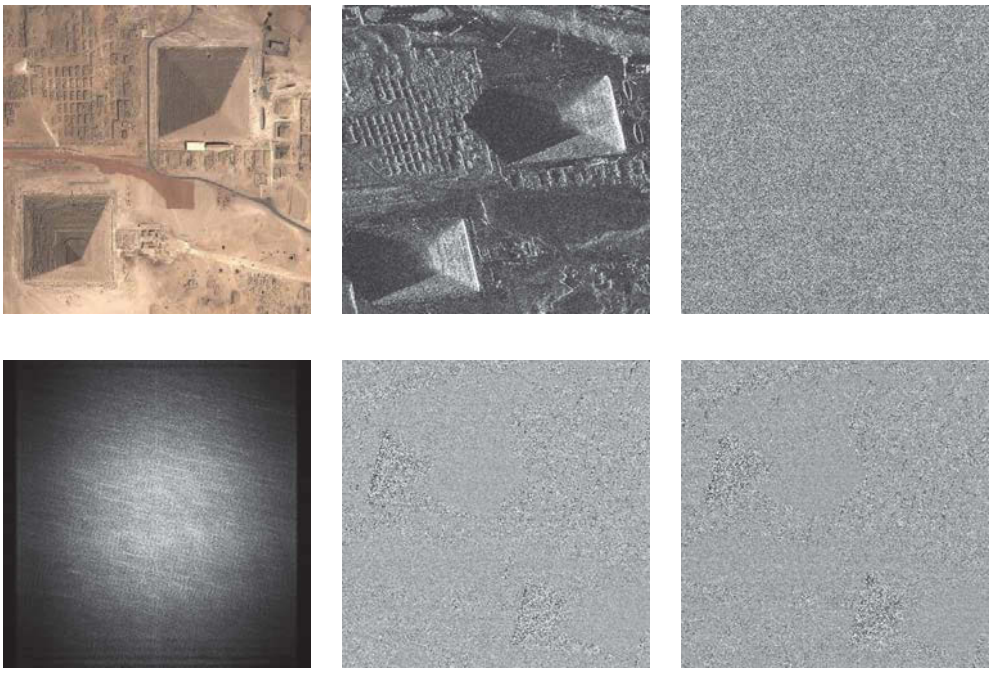


Figure 2.1: An optical satellite image versus a synthetic aperture radar (SAR) satellite image. Clockwise: Optical image for ground truth,© Google Inc.; amplitude, phase, real channel and imaginary channel of the complex-valued SAR image; amplitude-envelope of the complex-valued spectrum of SAR image. The SAR image has been taken from a full-scene of very-high-resolution spotlight mode of TerraSAR-X. Interesting scattering mechanisms taking place in this particular SAR image are demonstrated in Bamler & Eineder [2008].

Application areas of SAR includes surveillance, target detection, navigation, change detection, interferometric applications, environment monitoring and planning etc. The advent of high-resolution SAR has introduced a possibility of extensive target signature analysis, thus opening new windows in automatic target detection and recognition. To understand target's backscattering behavior, an understanding of SAR image formation plays a vital role. So, this chapter mainly comprises of the imaging system of SAR. An introduction to statistical perspective on SAR images is also presented.

2.1.1 Data acquisition geometry

A SAR sensor works by moving along its trajectory, transmitting coherent modulated electromagnetic pulses to the target in a side looking fashion, recording the backscattered signal and then performing signal processing operations to generate a final image. Figure 2.2 shows a simplified SAR data acquisition geometry and later the definitions of these geometry related terms are presented.

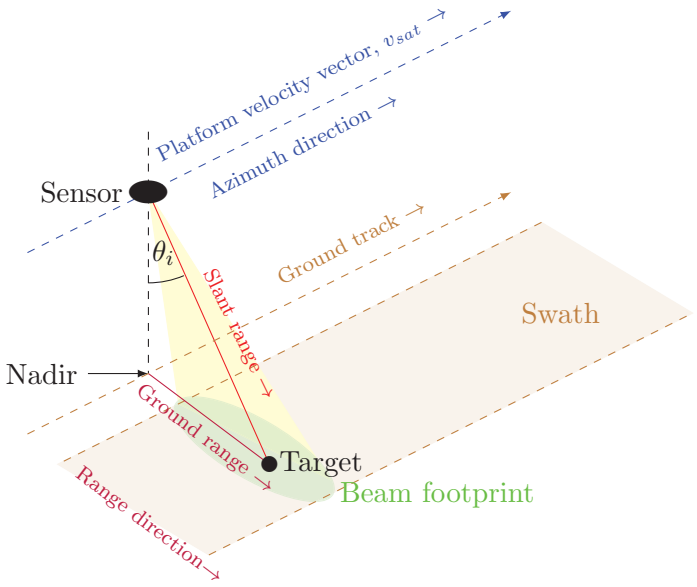


Figure 2.2: A simplified SAR data acquisition geometry along with the description of geometry related terms, some of which will be readily used in further discussions.

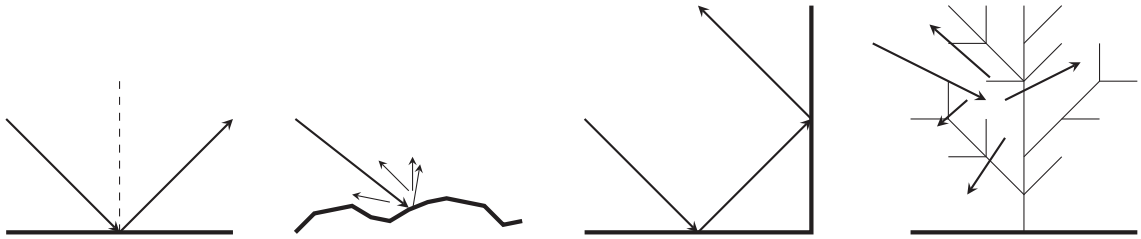
- **Sensor** The sensor is carried on a platform, which could be a satellite in the space-borne case or aircraft in the air-borne case, which moves along its trajectory transmitting electromagnetic pulses with side look angle, $\boldsymbol{\theta}_i$, and recording the echoes in a coherent manner.
- **Beam footprint** The area illuminated by the SAR sensor on the ground is called the beam footprint. The position and shape of the antenna footprints depends upon the antenna beam pattern as well as the geometry of the ground surface.
- **Target** Although the sensor illuminates an area on the ground, the target can be assumed as a hypothetical point on the ground under consideration, used to simplify mathematical computations.

- **Nadir** The nadir is the point directly below the sensor on ground. Nadir tracks the azimuth direction with the motion of platform.
- **Azimuth and range** The coordinate aligned with the velocity vector of the sensor is denoted by azimuth or *slow time*, whereas the across track coordinate is denoted by range or *fast time*.
- **Slant range and ground range** Slant range is the line joining the sensor and the target, ground range is the projection of slant range on the ground.

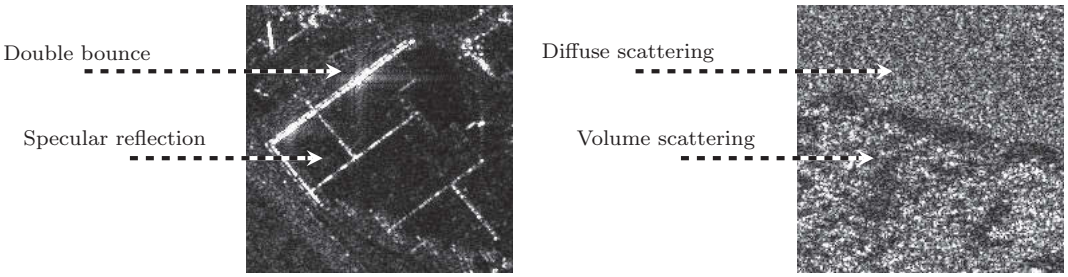
2.1.2 Geometrical perspective

2.1.2.1 Scattering mechanisms in SAR

Due to the side looking nature of the SAR sensor, several basic backscattering phenomena can be explained. Depending upon the microscopic and macroscopic properties of the scatterers on the imaged surface, some of the basic backscattering mechanisms in SAR are depicted in the figure 2.3(a).



(a) Left to right: Surface scattering on plane surface and rough surface, Double bouncing reflection, Volume scattering.



(b) Depiction of basic scattering mechanism on real SAR images.

Figure 2.3: Different backscattering mechanisms in SAR depending upon the microscopic and macroscopic properties of the scatterers on the imaged surface. First row from left to right: surface scattering (plane surface), surface scattering (rough surface), double bouncing reflection and volume scattering. Second row: depiction of these basic scattering mechanisms on metric-resolution SAR images from TerraSAR-X.

Surface scattering

Surface scattering mainly involves two phenomena. First is the surface scattering from smooth surfaces, also called the specular reflection. If the incidence angle between the transmitted SAR

signal and the imaged surface is not zero, very little energy is reflected back to the SAR antenna. Thus the imaged surface appears dark in the SAR image. That's why the flat rooftops of the buildings, calm water in the river or sea surfaces, surfaces on the free-ways appear dark in SAR images as very little energy is reflected back to the SAR antenna due to specular reflection. An example of specular reflection from the rooftop of a building is depicted in the figure 2.3b(left). The second phenomenon in surface scattering is scattering from the rough surfaces, also called the diffuse scattering. In this case only a part of the transmitted SAR signal energy is received back by the SAR antenna, such as from the crop-land, grass-fields, and sand dunes etc.. The imaged surface is brighter than the areas that show specular reflection behavior. An example of diffuse scattering from a grass-field is depicted in the figure 2.3b(right).

Single and double bouncing reflection

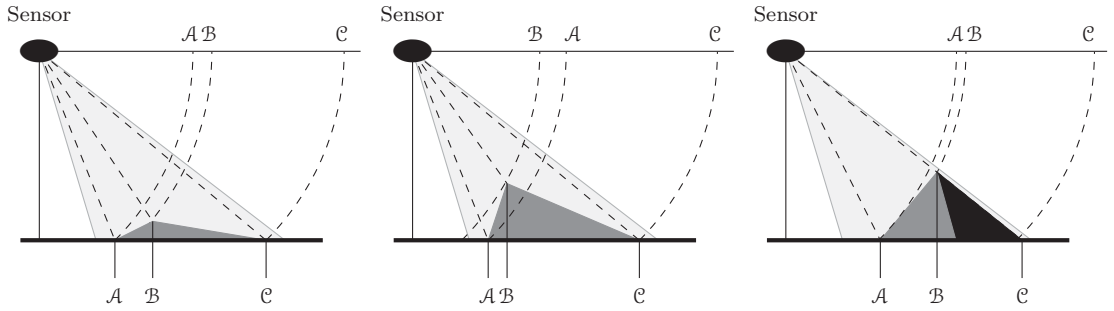
Single, double bounce reflection or multiple reflections occur from the edges or corners of strong structures (e.g. buildings, metal or concrete bridges etc.). That's why such objects appear very bright in SAR images due to strong backscatter from sharp edges on the borders. An example of double bounce from a wall of a building is depicted in the figure 2.3b(left).

Volume scattering

Volume scattering also occurs on the rough surfaces where the SAR signal partly penetrates the scatterers. The depth of penetration is a function of the employed wavelength in the SAR imaging signal. An example of volume scattering from the trees in a forest is depicted in the figure 2.3b(right).

2.1.2.2 Geometrical distortions in SAR

Again, due to the side-looking imaging geometry of SAR, some geometrical distortions are also intrinsic to SAR images depending upon the topography of the terrain (see figure 2.4). Here simplified flat Earth is considered ignoring the curvature of Earth. If θ_i is the look angle which increases from near range to far range, α is the local slope, then the local incidence angle becomes $\theta_i - \alpha$. Depending upon the relation between look angle and local slope, normally, the following three distortion cases are of interest (Franceschetti & Lanari [1999]):



(a) Foreshortening (b) Layover (c) Shadow

Figure 2.4: Geometrical effects in SAR images which arises due to the side-looking imaging geometry of the SAR sensor. These effects essentially depends upon the topography of the terrain.

Foreshortening $-\theta_i < \alpha < \theta_i$. The slant range distance $\mathcal{A}$-$\mathcal{B}$ is shorter than the distance on the ground, i.e. foreshortening causes dilation or compression of resolution cell.

Layover $\alpha \geq \theta_i$. The top of the mountain $\mathcal{B}$ is imaged ahead of the lower edge $\mathcal{A}$. Layover results in the inversion of the imaging geometry. In special case, when $\theta_i = \alpha$, the area is compressed on slant range into a single pixel.

Shadow $\alpha \leq \theta_i - \pi/2$. The region $\mathcal{B}$-$\mathcal{C}$ in the radar shadow is not imaged, so no backscattered signal is produced.

2.2 SAR signal processing

Radar equation establishes a physical correlation between power transmitted by the sensor and the power received from the target after backscattering. Considering the case of radar with real aperture, if P_t is transmitted power, and P_r is received power, then the following relation holds for unfocused systems:

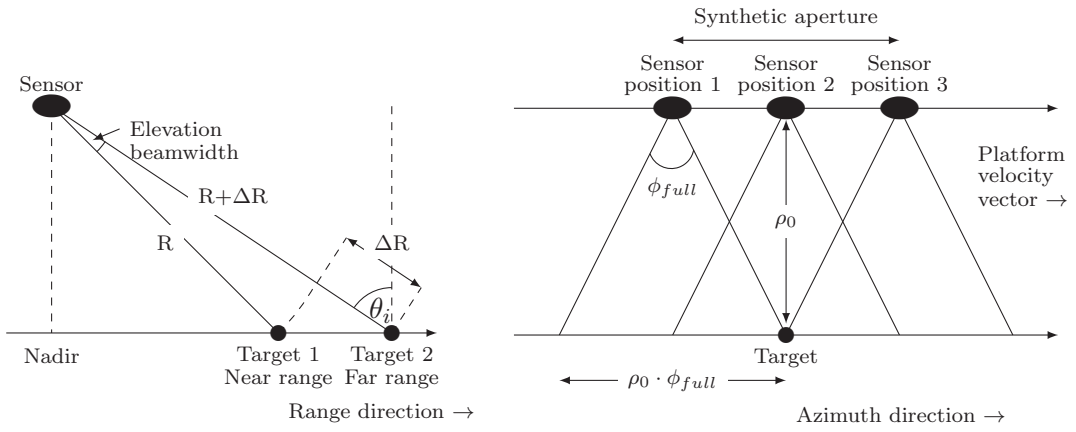
$$P_r = P_t \frac{\lambda^2}{4\pi} \sigma \frac{G^2}{(4\pi R^2)^2}, \tag{2.1}$$

where σ is the radar cross-section, a measure of energy backscattered per unit surface area by the target. The intensity observed in the radar image is proportional to σ (Curlander & McDonough [1992]). Here, λ is the carrier wavelength of the signal, R is the distance between target and sensor, G is the gain factor of antenna. It is assumed that transmission and reception gain, G_t and G_r respectively, of the antenna are same, and in the other case G^2 is replaced by $G_t G_r$ in the equation 2.1.

We consider again the approximated SAR data acquisition geometry shown in figure 2.2, where the curvature of Earth and curved flight trajectory have been ignored. These rectilinear geometry approximations are sufficient for mathematical formulations (Bamler & Schaettler [1993]). Further, a *Start-and-go approach* (Hein [2004]) assumes that the satellite position does not change between sending and receiving of an electromagnetic pulse for a target under consideration. The altitude and velocity of the trajectory are also to be assumed as constant during acquisition time, and atmospheric attenuation or refraction have also to be ignored. Under these assumptions, the SAR signal processing chain can be understood to explain the limiting SAR resolution, which can be further extended to the real scenario. SAR resolution is differentiated between range and azimuth resolution. SAR signal processing enables us to achieve a better resolution as compared to conventional radar.

2.2.1 Range resolution

To understand the focusing in range, we refer to the figure 2.5a. Lets assume that we want to resolve two point scatterers: Target-1 and Target-2 separated with a range ΔR. To get a perfect resolution, we need to transmit a delta function pulse such that the time separation between return pulses from these two target will be $\Delta t = 2\Delta R/c$. It will be possible to separate these two targets with a system with time-sampling finer than Δt. Unfortunately, it is not practical to transmit a infinite-energy pulse like delta function to achieve a better resolution. The other solution is to use a very short wave-train with a time window of length τ. In order to have no overlap between the return pulses from two targets, the time window of wave-train τ should be less than $2\Delta R/c$. But as the short wave-trains have very low energy, a little energy will be

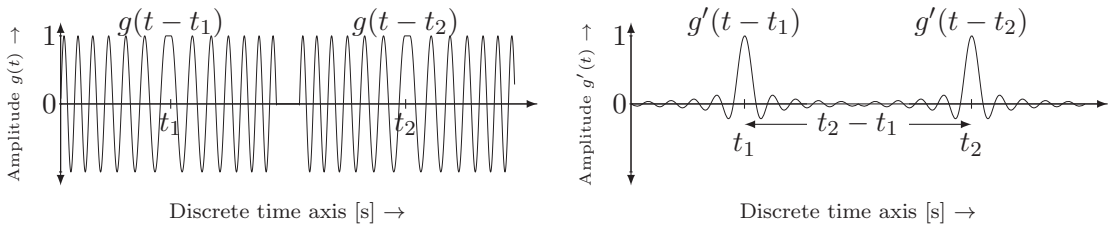


(a) Range resolution (b) Azimuth resolution

Figure 2.5: Resolution of SAR in range and azimuth directions. (a) SAR imaging geometry in range direction. (b) SAR imaging geometry in azimuth direction.

reflected from the targets. This little reflected energy will be suppressed with noise, thus makes the short wave-trains also impractical.

In order to circumvent this problem, the signal transmitted by the SAR sensor is a *chirp signal* with the form: $g(t) = \text{rect}(t/\text{T}) \cdot \exp(j\pi \text{K} t^2)$. It is an ideal linear frequency modulated signal, with time period T and linear frequency modulation rate (FM rate) K. The bandwidth (B.W.) of this chirp signal is described as the product of |K| and T. Again, in just transmitting the chirp and utilizing the returned chip signal without *pulse compression*, bandwidth governs the resolution in range direction. The signal is transmitted from the sensor towards the target, and the delayed backscatter is obtained with amplitude modified by the scattering coefficient of the target. Two targets in range can only be resolved, if the distance between them is more than T in time. Figure 2.6a displays backscattered signals from two targets in the non-compressed case.



(a) Non-compressed backscatter (b) Compressed backscatter

Figure 2.6: Non-compressed backscatter (after demodulation) and compressed backscatter using pulse compression technique, from two targets.

Pulse compression technique is used to improve the resolution in range. Pulse compression is achieved by matched filtering operation, i.e. taking the correlation between transmitted and received signal. A matched filter is used because it is the optimal linear filter in the sense of providing the best signal-to-noise ratio (Curlander & McDonough [1992]). Figure 2.6b displays range compressed signals received from two targets. In this case, it can be seen from the figure

that: $t_2 = t_1 + 1/\text{B.W.}$. So the range resolution in this case becomes:

$$\Delta R = R_{\text{slant-range}} = \frac{c}{2 \cdot |K| \cdot T}. \tag{2.2}$$

It can be seen from the equation 2.2 that the range resolution does not depend upon the distance between the SAR sensor and target. As the ground range is the projection of slant range on ground, slant range resolution can be converted into ground range resolution using the look angle information from the figure 2.5a, and is given by:

$$R_{\text{ground-range}} = \frac{c}{2 \cdot |K| \cdot T \cdot \sin(\theta_i)}. \tag{2.3}$$

This signal processing in range direction, also denoted as *focusing in range*, enables us to achieve a very good resolution in range direction. Further signal processing in azimuth direction is done to improve the resolution in azimuth direction.

2.2.2 Azimuth resolution and azimuth chirp

In case of real aperture radar, azimuth resolution is related to beam-width of the main lobe of antenna beam pattern. Figure 2.7 shows a typical beam pattern of a radar antenna with main lobe in the center. With the antenna length - L , and transmitted wavelength - λ, 3 dB-full

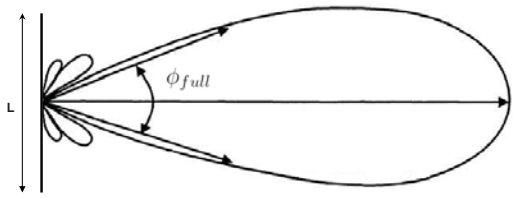


Figure 2.7: Antenna beam pattern.

width at half maximum of antenna's main lobe is given as: $\phi_{full} \approx 0.88 \cdot (\lambda/\text{L})$. In case of real aperture, azimuth resolution is dependent upon the distance of target from sensor. At a distance ρ_0, azimuth resolution is related to 3 dB-full width at half maximum as:

$$R_{\text{azimuth RAR}} = \phi_{full} \cdot \rho_0 = 0.88 \cdot \left(\frac{\lambda}{\text{L}}\right) \cdot \rho_0. \tag{2.4}$$

In order to achieve a good resolution, the physical length of the antenna needs to be increased, which has operational constraints. This gives an unacceptably poor value for azimuth resolution in the conventional radar case, i.e. the real aperture radar (RAR) case. It is evident from equation 2.4 that in order to improve azimuth resolution, either wavelength of the transmitted signal has to be reduced, or the size of antenna has to be increased. Wavelength of operation cannot be changed as the backscatter characteristics of targets depend upon the wavelength. Due to constructive limits, it is also not possible to increase the size of the antenna beyond certain limits, so the concept of synthetic aperture comes into play.

Figure 2.5b describes the perception of synthetic aperture in SAR. A target is continuously illuminated by the antenna beam during the trajectory of the sensor in azimuth direction over an extended period of time, known as *integration time*, in order to synthesize a larger size of an antenna. The backscatters of the target are stored into a sampling matrix along with the phase

information. The duration in which the beam is illuminating a target during its travel is known as synthetic aperture. Due to the change in distance between sensor and target during synthetic aperture, the echo signal shows a Doppler frequency shift in relation to the transmitted pulse. This Doppler shift stored in form of phase information is used later in signal processing and thus helps in achieving an improved azimuth resolution. To focus a SAR image in azimuth, the classical matched filter strategy is employed, i.e. the energy backscattered by the target is concentrated in time by a proper integration of Doppler phase history.

Assuming the satellite platform velocity is v_{sat}, if a target is on ground at range and azimuth coordinates $(\rho_0, v_{sat} \cdot (t - t_0))$, where t_0 is the start time, then its overall instantaneous distance from the sensor is given as:

$$\rho_{(\rho_0, t-t_0)} = \sqrt{\rho_0^2 + v_{sat}^2 \cdot (t - t_0)^2} = \rho_0 \left(\sqrt{1 + \frac{v_{sat}^2 \cdot (t - t_0)^2}{\rho_0^2}} \right). \tag{2.5}$$

We denote $(t - t_0) = t_{az}$ as the azimuth time. Approximating the equation 2.5 using Taylor series expansion, we can write:

$$\rho_{(\rho_0, t_{az})} \simeq \rho_0 + \frac{v_{sat}^2 \cdot t_{az}^2}{2\rho_0}. \tag{2.6}$$

For a particular instant in time, the difference between instantaneous range $\rho_{(\rho_0, t_{az})}$ and minimum range (or slant range) ρ_0 is called *range migration*, and can be written as:

$$\Delta\rho_{(\rho_0, t_{az})} = \rho_{(\rho_0, t_{az})} - \rho_0 \simeq \frac{v_{sat}^2 \cdot t_{az}^2}{2\rho_0}. \tag{2.7}$$

The range migration has the form of parabola as shown in the figure 2.8(left), thus a two-way path phase is introduced. The phase information extracted from this range variation follows

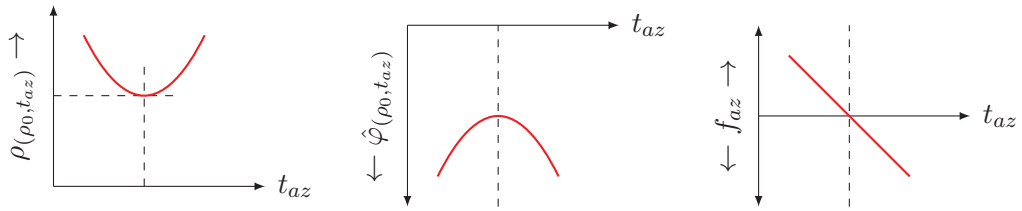


Figure 2.8: Concept of range migration. Left: the variation of the distance between the sensor and target $\rho_{(\rho_0,t_{az})}$ with respect to azimuth time t_{az}, Center: the generation of azimuth chirp as a quadratic phase modulation range migration $\hat{\varphi}_{(\rho_0,t_{az})}$, Right: a negative slope line for the linear frequency modulation rate or the instantaneous Doppler frequency f_{az} of the azimuth chirp.

approximately quadratic behavior, as shown in the figure 2.8(center). In other words, a linear frequency modulation is generated by the range migration, called the *azimuth chirp* with following form.

$$\hat{\varphi}_{(\rho_0, t_{az})} = -\frac{4\pi}{\lambda}\rho_{(\rho_0, t_{az})} \simeq -\frac{4\pi}{\lambda}\left(\rho_0 + \frac{v_{sat}^2 \cdot t_{az}^2}{2\rho_0}\right). \tag{2.8}$$

Now the *instantaneous Doppler frequency* f_{az} can be given as:

$$f_{az} = \frac{1}{2\pi}\frac{d}{dt}\hat{\varphi}(\rho_0, t_{az}) \simeq -\frac{2v_{sat}^2}{\lambda\rho_0}t_{az} = \mathrm{FM}_{\rho_0} t_{az}. \tag{2.9}$$

Here $\mathrm{FM}_{\rho_0} = -(2v_{sat}^2)/(\lambda\rho_0)$ is the frequency modulation rate of the azimuth chirp. As a consequence, the signal is a negative slope line with respect to azimuth frequency f_{az} versus azimuth time t_{az}, as shown in the figure 2.8(right). This instantaneous Doppler frequency at antenna beam center, i.e. $t_{az} = t_c$, is called *Doppler centroid* f_{DC}, and is denoted as:

$$f_{\mathrm{DC}} = f_{az}(t_{az} = t_c) = -\frac{2v_{sat}^2}{\lambda\rho_0}t_c. \tag{2.10}$$

In case of standard SAR mode (stripmap and no antenna squint), the Doppler centroid is at the center of the azimuth spectrum. For the spotlight mode case resulting central frequency is different for each one of the final azimuth lines in the Doppler coordinates, which needs to be taken into consideration for further applications.

The range migration has to be corrected before performing *focusing in azimuth*, but the Doppler history information available in range migration is also of significant importance for focusing in azimuth and the information available in range migration is required for focusing in azimuth, so focusing in range and azimuth is a coupled problem, and hence SAR imaging becomes a two-dimensional and a non-separable problem (Bamler & Schaettler [1993]).

Now we will derive the formulation for the resolution of SAR in azimuth. Going back to equation 2.9, the *instantaneous Doppler frequency* f_{az} can be rewritten as:

$$f_{az} \simeq -\frac{2v_{sat}}{\lambda\rho_0}v_{sat}t_{az}. \tag{2.11}$$

The magnitude of Doppler frequency for an short duration of time in azimuth Δt_{az} can be written as:

$$\mid \Delta f_{az} \mid \simeq \frac{2v_{sat}}{\lambda\rho_0}v_{sat}\Delta t_{az} = \frac{2v_{sat}}{\lambda\rho_0}\Delta\rho_{az}. \tag{2.12}$$

Here, $\Delta\rho_{az} = v_{sat}\Delta t_{az}$ is the shortest distance resolvable in the azimuth direction, and hence the azimuth resolution can be written as:

$$R_{\mathrm{azimuth}}(\text{ or } \Delta\rho_{az}) = \frac{\mid \Delta f_{az} \mid \rho_0\lambda}{2v_{sat}} \tag{2.13}$$

Now, the frequency resolution of the generated azimuth chirp signal is equivalent to the reciprocal value of time signal within the half-power bandwidth, i.e. $\mid \Delta f_{az} \mid = (0.88/\mathrm{T})$. The time period T is the duration of azimuth chirp which is related to the distance between the transmitted pulses S as $\mathrm{T} = \mathrm{S}/v_{sat}$. We recall the distance S between transmitted pulses from the equation 2.4 as $\mathrm{S} = \phi_{full} \cdot \rho_0$, which is also the resolution for a real aperture radar. Now the frequency resolution of the generated azimuth chirp signal can be written as follows (Hein [2004]):

$$\mid \Delta f_{az} \mid = \frac{0.88}{\mathrm{T}} = \frac{0.88}{\mathrm{S}/v_{sat}} = \frac{0.88\ v_{sat}}{0.88\left(\frac{\lambda}{\mathrm{L}}\right)\rho_0}. \tag{2.14}$$

Substituting the equation 2.14 in the equation 2.12, we obtain the azimuth resolution in SAR as:

$$R_{\mathrm{azimuth}}(\text{ or } \Delta\rho_{az}) = \frac{\mathrm{L}}{2}. \tag{2.15}$$

It is evident again from the equation 2.15 that azimuth resolution is independent of antenna distance from sensor.

Concept of azimuth spectrum splitting in SAR

After the discussion on azimuth chirp, it is pertinent to briefly discuss the concept of azimuth splitting. The position and frequency information is interrelated in the SAR phase history, which is utilized in the concept of azimuth spectrum splitting. Azimuth spectrum splitting (or azimuth sub-band/sub-aperture in the complex spectrum of the SAR signal) corresponds to the splitting of the full synthetic aperture in order to extract a specific viewing angle of the sensor. Each Doppler phase term in the Doppler effect generated due to the radar-target range variations during the SAR acquisition is related to a specific sensor position and hence to a specific azimuth viewing angle. In the figure 2.9, a pictorial representation concept of azimuth splitting by extracting 4 sub-apertures from the full synthetic aperture is demonstrated.

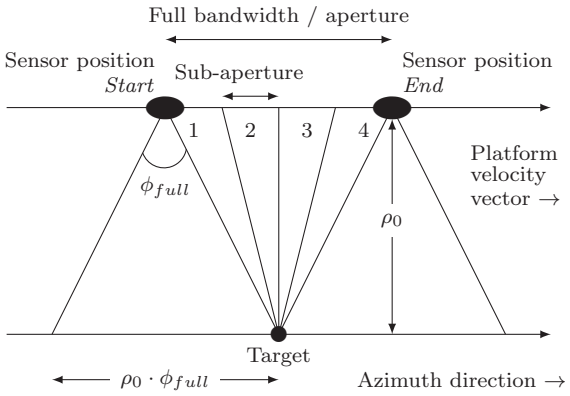
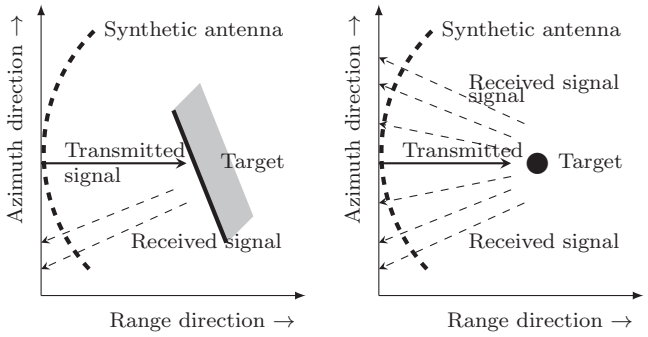


Figure 2.9: Concept of extracting azimuth sub apertures from the complex-valued spectrum. The extraction of 4 non-overlapping consecutive sub apertures is demonstrated. The full bandwidth is the consequence of the full synthetic aperture generated during the SAR image acquisition from Sensor position *Start* to Sensor position *End*, where a particular target is visible in the half-power bandwidth of the main lobe of the antenna beam.

Extracting the part of this spectrum is equivalent to selecting a sub-antenna with an particular azimuth squint angle, thus providing the information about the directivity of the objects, their orientation and the material of the surrounding surfaces etc. Illustration of the backscattering direction for a diedre and a coherent scatterer is presented in the figures 2.10a and 2.10b respectively.



(a) Non-coherent target (b) Coherent target

Figure 2.10: Backscattering from two different targets, (a) a diedre and (b) a coherent scatterer.

In the figure 2.10a, the backscatters from a highly directive and non-coherent object will not be visible in all of the sub-apertures of the synthetic antenna, whereas in figure 2.10b, the backscatters from a coherent object will be visible in all sub-apertures. This phenomenon, which is peculiar to the complex azimuth spectrum can also be extended to the range direction to reveal some otherwise hidden information about the objects on ground.

Azimuth ambiguities

Azimuth ambiguities in SAR are caused by the aliasing of the Doppler phase history of each target that is sampled according to the sensor azimuth sampling frequency, which equals the pulse repetition frequency (PRF) (Li & Johnson [1983]). We refer to the figure 2.11 in order to understand azimuth ambiguities in SAR images.

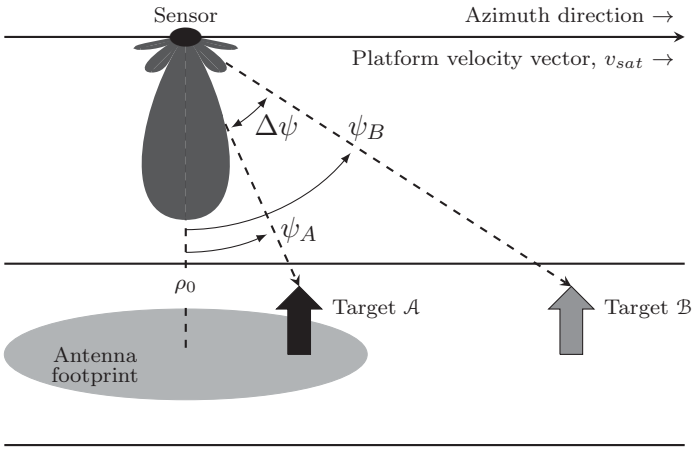


Figure 2.11: Azimuth ambiguities in SAR images. The target $\mathcal{B}$ will be aliased with the target $\mathcal{A}$ if the angular azimuth frequency corresponding to the angular displacement $\Delta\psi$ is a multiple of the sampling frequency (Guarnieri [2005]).

The satellite is flying with velocity v_{sat} on a straight path, ignoring the curvature of the Earth for time being, and two targets under consideration on a constant line of range are target $\mathcal{A}$ and target $\mathcal{B}$, with respective azimuth angle to be ψ_A and ψ_B. The contribution of the target $\mathcal{B}$ will be displaced by an angular displacement of $\Delta\psi = \psi_B - \psi_A$ with respect to target $\mathcal{B}$. We recall the formulation for the azimuth frequency $f_{az} = -(2v_{sat}/\lambda)(v_{sat}t_{az}/\rho_0)$ from the equation 2.9. Now for a small bandwidth SAR and a smaller angle $\Delta\psi$, the term $(v_{sat}t_{az}/\rho_0)$ can be approximated as $\sin\Delta\psi$ or simply $\Delta\psi$. Thus we have the relation between the azimuth frequency and the angle as follows:

$$f_{az} \simeq -\frac{2v_{sat}}{\lambda\rho_0}v_{sat}t_{az} \simeq -\frac{2v_{sat}}{\lambda}\sin\Delta\psi \simeq -\frac{2v_{sat}}{\lambda}\Delta\psi. \tag{2.16}$$

The two contributions from the target $\mathcal{A}$ and target $\mathcal{B}$ become indistinguishable when the azimuth frequency corresponding to the angular displacement $\Delta\psi$ is a multiple of the sampling frequency $f_{sa} = \text{PRF}$ (Guarnieri [2005]):

$$\frac{2\Delta\psi}{\lambda} = -\frac{f_{sa}}{v_{sat}} \Rightarrow \Delta\psi = -\frac{\lambda}{2v_{sat}}f_{sa}. \tag{2.17}$$

These aliased contribution from $\mathcal{B}$ also have the quadratic phase history. After azimuth focusing, these targets are somehow focused but have ghosts at the wrong position.

2.2.3 Spotlight mode SAR and concept of Doppler phase de-ramping

For a conventional stripmode SAR, the theoretical possible azimuth resolution R_{azimuth} is range-independent and depends upon the size of antenna L with $R_{\text{azimuth}} = \text{L}/2$. The azimuth resolution cannot be increased arbitrarily by reducing the size of antenna L, as this would result in the reduction in swath width and an increase in the risk of range ambiguities. This limiting factor on azimuth resolution is overcome in spotlight mode SAR by sweeping the azimuth beam backward during the imaging, by either mechanically rotating the SAR sensor (as in German military satellite SAR-Lupe), or by steering a phase-array antenna electronically (as in TerraSAR-X, and COSMO-SkyMed SAR satellites). A pictorial representation of sweeping the azimuth beam backward is shown in the figure 2.12.

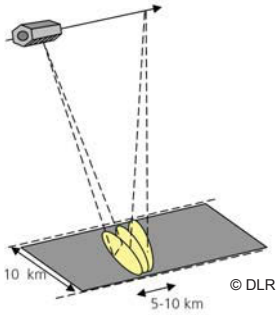


Figure 2.12: Spotlight mode imaging geometry for TerraSAR-X.

In the figure 2.13(left) is shown the case of conventional stripmap.

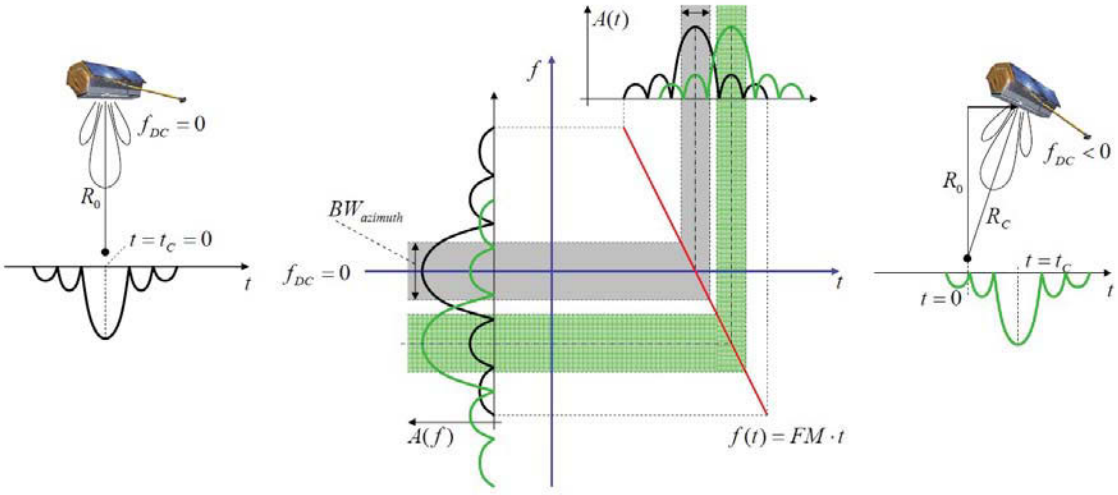


Figure 2.13: Effect of squint of the antenna onto the acquired Doppler frequencies f_{az}. As described by the green area, the squinting of antenna backwards will result in the drift of the Doppler centroid frequency. Thus the Doppler centroid in the generated Doppler chirp will not be centered in contrast to the no-squint antenna, as in the figure 2.8 (right). Image source: Eineder *et al.* [2008].

Now in order to increase the resolution in azimuth, the illumination of the target is extended in time by modifying the squint angle during acquisition, as shown in figure 2.13(right). With this non-stationary squint angle, the azimuth chirp is acquired for a longer duration. On proper integration this provides better resolution in azimuth. However, this non-stationary squint angle

in spotlight mode SAR imaging results in a negative drift of the Doppler centroid frequency in images, as shown by the green area in the figure 2.13(center). This negative drift of the Doppler centroid frequency, i.e. $f_{\mathrm{DC}} < 0$, needs to be considered while dealing with the complex-valued SAR images from the spotlight mode acquisition. The maximum illumination time for a synthetic aperture in spotlight mode SAR is given as:

$$t_{\mathrm{aperture}} = \frac{\mathrm{B.W.}}{\dot{f}_{az} - \mathrm{FM}}. \tag{2.18}$$

Here, $\dot{f}_{az}$ is the Doppler rate caused by sweeping the antenna in the backward direction, B.W. is the antenna Doppler bandwidth, and FM is the frequency modulation rate. Thus, the focused-image azimuth time interval $\Delta t_{\mathrm{focused}}$ will be shorter than the recorded raw data time interval Δt_{RAW} as shown in the figure 2.14(left). The relation between $\Delta t_{\mathrm{focused}}$ and Δt_{RAW} can be written as follows (Eineder *et al.* [2009]):

$$\Delta t_{\mathrm{focused}} = \Delta t_{\mathrm{RAW}} + \frac{\mathrm{B.W.} - \Delta t_{\mathrm{RAW}} \dot{f}_{az}}{\mathrm{FM}}. \tag{2.19}$$

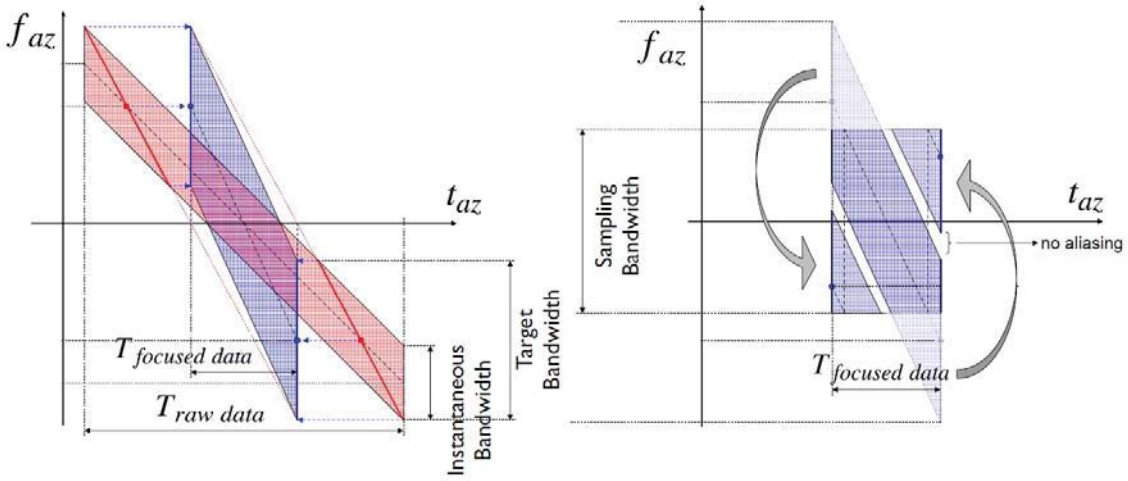


Figure 2.14: Focusing (left) and sub sampling (right) of the SAR spotlight product. The negative drift of the Doppler centroid frequency of the slope in the azimuth frequency f_{az} with respect to the azimuth time t_{az} represents the Doppler rate $\dot{f}_{az}$ caused by the sweeping the antenna. Image source: Eineder *et al.* [2008].

Now the bandwidth B.W.$_{\mathrm{spotlight}}$ in the case of spotlight mode SAR as a product of improved aperture time t_{aperture} and FM can be expressed as follows:

$$\mathrm{B.W.}_{\mathrm{spotlight}} = t_{\mathrm{aperture}}\mathrm{FM} = \frac{\mathrm{B.W.}}{\dot{f}_{az} - \mathrm{FM}}\mathrm{FM} = \frac{\mathrm{B.W.}}{\dot{f}_{az}/\mathrm{FM} - 1}. \tag{2.20}$$

We notice that the bandwidth B.W.$_{\mathrm{spotlight}}$ in spotlight mode is higher than the antenna Doppler bandwidth B.W., which require sampling the image with a higher frequency than the PRF of the raw data while focusing. Equations 2.18- 2.20 will describe the conventional stripmap mode without squint angle, where there is no sweeping of the antenna resulting in $\dot{f}_{az} = 0$.

Even after focusing the raw data to the single look complex (SLC) images in the spotlight mode SAR, systematic Doppler centroid drift remains as shown in the figure 2.14(right). To

remove the linear negative Doppler drift in the focused scene and to center the spectrum around the start Doppler frequency, a phase de-ramping algorithm must be performed prior to other operations on the complex-valued SAR images, by multiplying each azimuth line with a chirp function parametrized with the Doppler drift rate (Eineder *et al.* [2009]):

$$c(t) = \exp(-j\pi(t - t_{\text{start}})^2 \dot{f}_{az}). \tag{2.21}$$

Where, t_{start} is the zero-Doppler time of the first azimuth sample. All the SLC spotlight mode products from TerraSAR-X used in this thesis have been first phase de-ramped using the equation 2.21. Essential Doppler centroid parameters needed for phase de-ramping are available in the XML-formatted TerraSAR-X description file for each product.

2.3 Statistical analysis of SAR images

A SAR system works on the principle of transmitting a coherent modulated electromagnetic pulse to the target in a side looking fashion, recording the backscattered signal and then performing signal processing operations to generate a final image. The raw SAR image or the signal received from the target is inherently a complex-valued signal. Even after focusing to obtain the final image, the SAR images are complex-valued, i.e. it constitutes a magnitude and a phase term. However for many application, this phase information is ignored and only magnitude term is utilized. In order to derive some statistical properties of SAR images, we need to consider the complex-valued SAR signals.

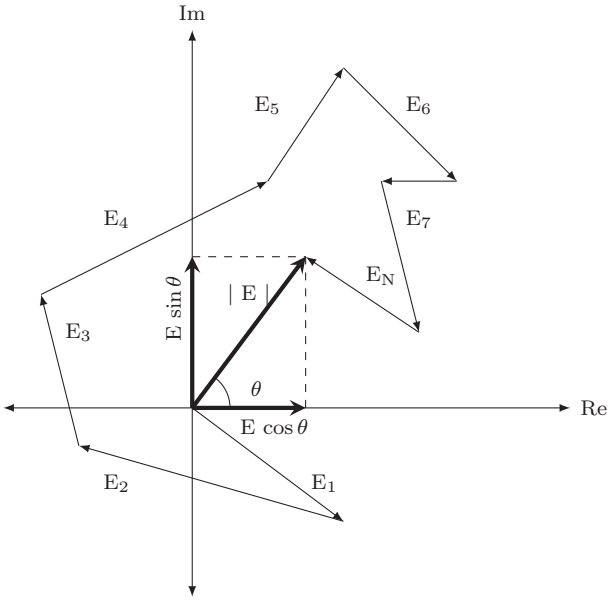


Figure 2.15: Random walk in a complex plane. The complex return from a resolution cell in the imaged surface is the superposition of **N** number of contributing backscattered waves with random phases.

The objective of statistical modeling is to address the problem of constructing a stochastic model to predict the behavior of a random process. To explain a standard model for the SAR images, a random walk phenomenon in the complex plane is employed. According to this hypothesis, a resolution cell consists of **N** number of discrete scatterers. The resulting backscattered complex signal **E** in a random walk is interpreted as the sum of all **N** individual

scatterers, as shown in the figure 2.15. In particular, this resultant is the contribution of the interference of the returns from corresponding **N** scatterers and this interference phenomenon motivates the usual noise-like granular aspect of SAR images, known as *speckle* (Oliver & Quegan [1998]). This phenomenon of speckle essentially affects all coherent imaging systems, such as laser (Goodman [1976]), acoustic and radar images.

If each scatterer provides a backscatter with magnitude E_i and phase term θ_i, the total returned modulation of the incident wave is given as:

$$\mid \mathbf{E} \mid \mathrm{e}^{j\theta} = \sum_{i=1}^{\mathbf{N}} E_i \mathrm{e}^{j\theta} \tag{2.22}$$

In a standard model for the statistical analysis of the complex signal, several assumptions are made. E.g. the number of scatterers in a resolution cell are large, the scatterers are independent and small (as compared with the ground area), the scattering instantaneous phases are independent of the amplitudes and uniformly distributed in $[0, 2\pi]$, and there is no single scatterer dominating the scene (Touzi [2002], Oliver & Quegan [1998]). Statistically, according to the central limit theorem, if the number of scatterers **N** is large, in-phase and quadrature components, i.e. the real and imaginary parts of **E** turn out to be independent zero-mean Gaussian random variables with equal variances (Papoulis [1991]). Under the circumstances that all these assumptions are met, which is the case of fully developed speckle, the statistical modeling of the SAR images is rather easy. However in real scenarios, the situation is not that simple. In this section we will consider the case of fully developed speckle for statistical understanding of SAR images.

2.3.1 Statistics for amplitude, phase and intensity

With real and imaginary parts as zero-mean Gaussian random variables, it is straightforward to derive the probability density functions (pdfs) for amplitude and intensity for **E**, i.e. for SAR images. Lets denote the complex-valued SAR image pixel as:

$$y = y_R + jy_I = r\mathrm{e}^{j\theta} = \sqrt{v}\mathrm{e}^{j\theta}, \tag{2.23}$$

with, y_R, y_I, r, v and θ as its real component, imaginary component, the amplitude, the intensity, and the phase of the complex signal y, respectively. The joint probability density function (pdf) for real and imaginary components is given as:

$$p_{y_R y_I}(y_R y_I) = p_{y_R}(y_R) p_{y_I}(y_I) = \frac{1}{\pi\sigma} \exp\left\{-\frac{{y_R}^2 + {y_I}^2}{\sigma}\right\}, \quad (y_R, y_I) \in \mathbb{R}^2, \tag{2.24}$$

which is the circular Gaussian distribution, with variance σ. The observed phase θ is uniformly distributed in $[0, 2\pi]$. The corresponding marginal amplitude pdf for amplitude $r = \sqrt{{y_R}^2 + {y_I}^2}$ can be written as:

$$p_r(r) = \int_0^{2\pi} p_{r\theta}(r, \theta) d\theta = \frac{2r}{\sigma} \exp\left\{-\frac{r^2}{\sigma}\right\}, \quad r \geq 0, \tag{2.25}$$

which is a Rayleigh distribution with its expected value $\mathbb{E}[r] = \sqrt{\pi\sigma}/2$ and standard deviation: $\sigma\sqrt{1 - \pi/4}$. The intensity v, which is the square of amplitude r exhibits an exponential distribution as follows:

$$p_v(v) = \frac{1}{\sigma} \exp\left\{-\frac{v}{\sigma}\right\}, \quad v \geq 0, \tag{2.26}$$

with mean value and standard deviation equal to σ. The distribution for log-amplitude SAR images follows the Fisher-Tippet distribution. Figure 2.16 presents the experimentally derived pdf in terms of its normalized histogram for a complex-valued SAR image from high-resolution spotlight mode of TerraSAR-X which has been shown in the figure 2.1.

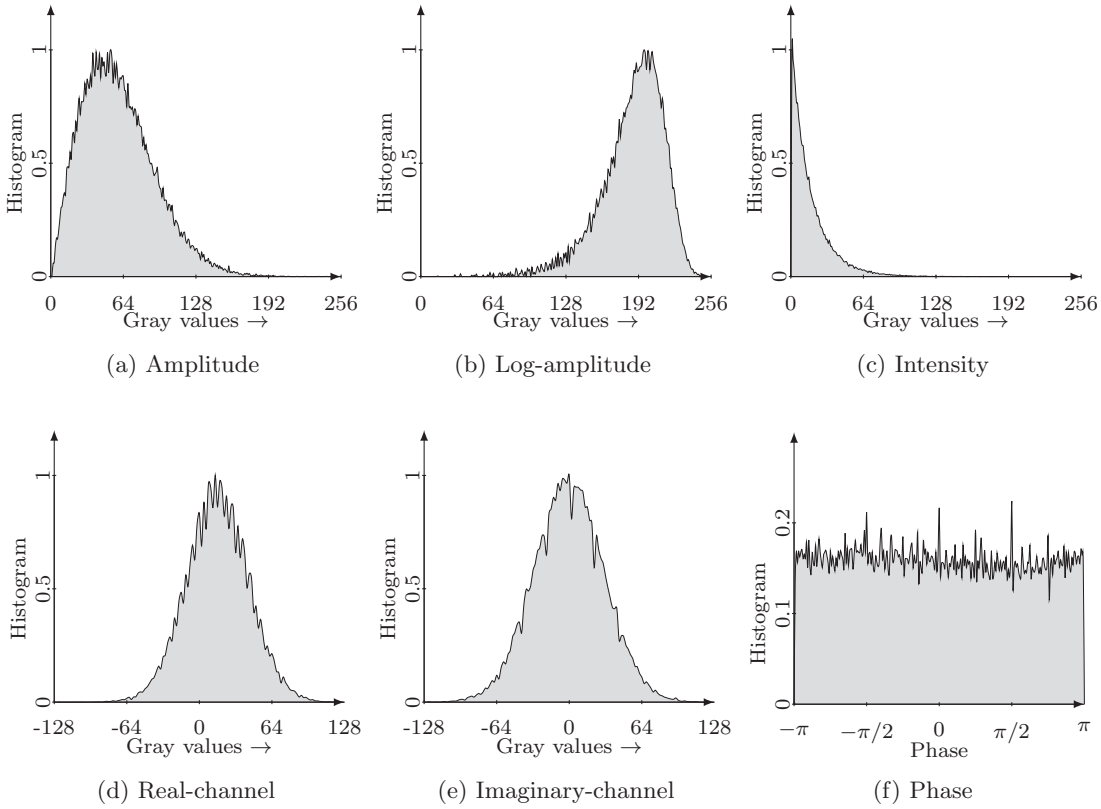


Figure 2.16: SAR image distribution for a complex-valued SAR image (very-high-resolution spotlight mode of TerraSAR-X). Normalized histogram of experimental distribution for (a) amplitude (Rayleigh distribution), (b) log-amplitude (Fisher-Tippet distribution), (c) intensity (Negative exponential distribution), (d) real channel (Gaussian distribution), (e) imaginary channel (Gaussian distribution), and (f) phase (uniform distribution in the interval $[-\pi, \pi]$).

The mentioned pdfs are of fundamental importance in handling SAR data, but are valid only for single-look SAR images with homogeneous areas. As multilook SAR images are used widely in practice, the Gamma distribution has been introduced as a model for a multilook SAR intensity pdf. The Gamma distribution is a generalization of the exponential distribution by averaging single-look exponentially distributed intensities, written as (Oliver & Quegan [1998]):

$$p_v(v) = \frac{(\lambda L)^L}{\Gamma(L)} v^{L-1} \mathrm{e}^{\lambda L v}, \quad v \geq 0; \tag{2.27}$$

with L as the equivalent number of looks (ENL) defined as the number of independent intensity values averaged per pixel (Oliver & Quegan [1998]), $\Gamma(\cdot)$ is the standard gamma function (Rudin [1976]), and λ is a distribution parameter. This results into a Nakagami distribution for a multilook SAR amplitude images, written as follows (Oliver & Quegan [1998], Tison *et al.*

[2004]):

$$p_r(r) = \frac{2(\lambda L)^L}{\Gamma(L)} r^{2L-1} \mathrm{e}^{-\lambda L r^2}, \quad r \geq 0. \tag{2.28}$$

With again, L as the equivalent number of looks (ENL) and $\Gamma(\cdot)$ is the standard gamma function.

More accurate pdf characterization

Coming back to the modeling of amplitude SAR images, where the assumption is that the real component y_R and the imaginary component y_I are zero mean Gaussian distributed. However in practice, in SAR images, the real component y_R and the imaginary component y_I often present significantly non-Gaussian behavior. Thus the amplitude SAR data r is also deviated from the Rayleigh empirical distribution. This issue necessitate a more accurate pdf characterization. The non-Gaussian behavior of y_R and y_I has been characterized by the symmetric α-stable distribution in Kuruoglu & Zerubia [2004], leading to a heavy-tailed Rayleigh model of amplitude SAR images:

$$p_r(r) = r \int_0^{+\infty} \rho \mathrm{e}^{-\gamma \rho^\alpha} J_0(r\rho) d\rho, \quad r \geq 0. \tag{2.29}$$

With, α, γ as the distribution parameters, and $J_0(\cdot)$ the zero-th order Bessel function of the first kind. The non-Gaussian behavior of y_R and y_I characterized by generalized Gaussian distribution in Moser *et al.* [2006] results in a generalized Gaussian Rayleigh (GGR) model for amplitude SAR images:

$$p_r(r) = \frac{\gamma^2 r}{\lambda^2 \Gamma^2(\lambda)} \int_0^{\pi/2} \exp[-(\gamma r)^{1/\lambda}(|\cos\theta|^{1/\lambda} + |\sin\theta|^{1/\lambda})] d\theta, \quad \lambda, \gamma > 0; r \geq 0. \tag{2.30}$$

With, λ, γ as the distribution parameters, and $\Gamma(\cdot)$ the standard gamma function. In the same way, with the help of two-sided generalized Gamma distribution, a generalized Gamma Rayleigh has been derived in Li *et al.* [2010]:

$$\begin{aligned} p_r(r) = \left[\frac{\nu}{\eta^{\kappa\nu}\Gamma(\kappa)}\right]^2 r^{2\kappa\nu-1} \int_0^{\pi/2} |\cos\theta \sin\theta|^{\kappa\nu-1} \\ \times \exp\left\{-\left[\frac{r}{\eta}\right]^\nu (|\cos\theta|^\nu + |\sin\theta|^\nu)\right\} d\theta, \quad r \geq 0. \end{aligned} \tag{2.31}$$

With, ν, κ, η as the distribution parameters.

In addition to this, several empirical models, without making any assumptions of the SAR physical concept and mechanism have also been reported for SAR images with a certain type of heterogeneous areas. For example, the well-known log-normal, Weibull, Fisher distributions etc. These models are obtained by straightforward induction of the experimental analysis on actual SAR images.

The log-normal distribution, with m and σ as distribution parameters, is written as follows (Oliver & Quegan [1998], Szajnowski [1977]):

$$p_r(r) = \frac{1}{\sigma r \sqrt{2\pi}} \exp\left[-\frac{(\ln r - m)^2}{2\sigma^2}\right], \quad m \in \mathbb{R}, \sigma > 0; r > 0. \tag{2.32}$$

The Weibull distribution, with η and μ as distribution parameters, can be written as follows (Oliver [1993], Oliver & Quegan [1998]):

$$p_r(r) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp\left[-\left(\frac{r}{\eta}\right)^\eta\right], \quad \eta, \mu > 0; r \geq 0. \tag{2.33}$$

The Fisher distribution, with $\mathcal{M}$ and $\mathcal{L}$ as the form parameters and μ as the mean, can be written as follows (Tison *et al.* [2004]):

$$p_r(r) = \frac{2\sqrt{\mathcal{L}}\Gamma(\mathcal{L}+\mathcal{M})}{\mu\sqrt{\mathcal{M}}\Gamma(\mathcal{L})\Gamma(\mathcal{M}))} \frac{\left[\sqrt{\frac{\mathcal{L}}{\mathcal{M}}}\frac{r}{\mu}\right]^{2\mathcal{L}-1}}{\left[1+\left(\sqrt{\frac{\mathcal{L}}{\mathcal{M}}}\frac{r}{\mu}\right)^2\right]^{\mathcal{L}+\mathcal{M}}}, \quad \mathcal{M} > \mathcal{L}; r \geq 0. \tag{2.34}$$

An alternative approach, to derive the theoretical parametric models for SAR images, is based on the multiplicative model. The multiplicative model is a common framework used to explain the stochastic behavior of data obtained with coherent illumination. The assumption made is that the observed pixel y in a SAR images is an outcome of the product of two independent random variables: one modeling the terrain backscatter x, which is generally considered real and positive, and other modeling the speckle noise z, which could be complex (for complex-valued SAR images) or positive real (for amplitude and intensity SAR images). So again with the assumption of a fully developed speckle, the multiplicative noise model for SAR images is written as follows (Touzi [2002]):

$$y = x \cdot z. \tag{2.35}$$

With, y as the observed image pixel, x is the noise free pixel, i.e. the backscatter information, and z is the speckle noise. Most common models based on multiplicative noise model are $\mathcal{K}$ and $\mathcal{G}$ distributions.

In the $\mathcal{K}$ distribution (Oliver [1984]), both the underlying backscatter and speckle are supposed to have Gamma distributions, which lead to the following distribution for the amplitude, with $\mathcal{L}$, ν and μ as distribution parameters:

$$p_r(r) = \frac{2}{r}\left(\frac{\mathcal{L}\nu r}{\mu}\right)^{\frac{\mathcal{L}+\nu}{2}} \frac{1}{\Gamma(\mathcal{L})\Gamma(\nu)} K_{\nu-\mathcal{L}}\left(\sqrt{\frac{\mathcal{L}\nu r}{\mu}}\right), \quad r \geq 0. \tag{2.36}$$

where K is a modified Bessel function of the second kind. Here the backscatter follows a Gamma distribution with mean 1 and shape parameter $\mathcal{L}$. The speckle follows a Gamma distribution with mean μ and shape parameter ν.

A more generalized formulation for $\mathcal{K}$ distribution was presented in Frery *et al.* [1997], termed as $\mathcal{G}$ distribution. In this distribution, the backscatter is modeled as a generalized inverse Gaussian law. The density function of amplitude SAR images for $\mathcal{G}$ has the following form, with α, γ and n as distribution parameters:

$$p_r(r) = \frac{\Gamma(n)}{\Gamma(n)K_\alpha(2\sqrt{\lambda\gamma})} r^{n-1}\left(\frac{\gamma+nr^2}{\lambda}\right)^{(\alpha-n)/2} K_{\alpha-n}(2\sqrt{\lambda(\gamma+nr^2)}), \quad \alpha,\gamma \geq 0, n > 0; r \geq 0. \tag{2.37}$$

The objective behind this short discussion in this section was to present an overview on the state-of-the-art research efforts on modeling of the amplitude SAR images. However, it has been demonstrated that a single density distribution can not characterize SAR images in all the possible scenarios (Singh *et al.* [2012]).

2.4 Summary

The following points have been discussed in this chapter.

- SAR imaging systems work on the principle of transmitting coherent modulated electromagnetic pulses to a target in a side looking fashion and recording the backscattered signal to generate a final image after signal processing. Due to the side looking geometry, all the SAR images are affected by some distortions such as foreshortening, layover and/or shadow.
- The signal processing techniques such as pulse compression and synthetic aperture are performed to achieve an adequately higher resolution. The formation of synthetic aperture not only provides a better resolution in azimuth direction, but also leads to the development of the concept of azimuth splitting which provides directivity of objects and the material of the surrounding surfaces by looking into various sub-apertures in azimuth.
- To achieve even higher resolution in spotlight mode SAR, sweeping of the azimuth beam backwards during the imaging is carried out. Though it provides better resolution, it also introduces a bias in the azimuth phase which needs to be de-ramped before further processing of complex-valued images. Guidelines for the Doppler phase de-ramping have been presented.
- In the statistical analysis of SAR images, the phenomenon of speckle is of prime importance. Speckle is a multiplicative noise-like granular aspect in SAR images, which leads to simpler SAR statistical models such as Rayleigh and Negative exponential for amplitude and intensity images respectively.
- Going from simpler models for SAR images, many advanced and more general models for amplitude images have been proposed in literature such as generalized Gaussian Rayleigh, $\mathcal{K}$ distribution, $\mathcal{G}$ distribution and so on. Normally even in advanced statistical models, the assumption on the multiplicative nature of speckle is still preserved. However, its difficult to arrive at a consensus on the density model which best fits the SAR images in all scenarios.

After presenting the above introductory concepts of SAR images, in next chapter, chapter 3, we will discuss the more specific issue of high-resolution SAR images. The main contributions of this thesis will be highlighted.

Chapter 3

SAR Image Content

The objectives of this chapter are to discuss the problem of spatial content understanding in low-resolution and high-resolution SAR images, and to highlight the main contributions of this thesis. In the first section we discuss the issue of ever increasing resolution from the recent SAR satellite missions. The image content in the low and medium resolution SAR images could be described in terms of global classes by characterizing the underlying texture by the assumption of fully developed speckle. We present the examples of the high-resolution SAR images, where, apart from the visibility of texture for homogeneous regions, it is possible to analyze the objects in more detail.

As the information available in high-resolution SAR images is sometimes sufficient for individual target analysis, in the second section we propose a method of visual exploration of targets by extending the azimuth splitting technique, which we term as the multiple sub-look decomposition (MSLD) method.

In the third section, we introduce the problem of spatial content understanding which could be carried out using time-frequency analysis (TFA) methods or statistical methods. Based on the discussion from MSLD and TFA, we will analyze a proposed method for information extraction using the fractional Fourier transform in the following chapters. The statistical methods are mainly based on the multiplicative speckle model which leads to the Bayesian inference applied to SAR images. Thus we discuss a historical overview of speckle and speckle reduction filters to the Bayesian information extraction.

In the fourth section, we discuss the issue of pixel-based classification and patch-oriented image categorization, which leads us to a proposed methodological classification for performance analysis of various methods. The methods, analysis and results will be discussed in context to the proposed methodological classification in the coming chapters. We will close this chapter with a short summary in the last section.

3.1 SAR images: from low-resolution to metric-resolution

Seasat launched by USA in 1978, and maintained by JPL, was the first civilian Earth-orbiting satellite with an on-board SAR sensor. The objective behind the design of Seasat was to test various oceanographic sensors and demonstrate the feasibility of the monitoring of oceanographic phenomena to gain a better understanding of the Earth's seas. But this satellite had to be terminated due to malfunctioning in its operation. Since then, a number of SAR satellites have been launched. The choice of frequency band in which these satellites were/are operating is closely related to the application area of the received image data. For example, American SIR-C SAR satellite in L-Band (24.2 cm wavelength) and C-band (5.7 cm wavelength) could penetrate to the forest floor and made moisture measurements possible. The short wavelength of 31 mm in TerraSAR-X provides very-high-resolution of 1-meter on ground, which makes target detection and recognition possible. Future satellite missions such as Sentinel-1 are also aimed at providing enhanced revisit and coverage for surveillance, mapping and monitoring purposes, at the cost of lower resolution, which provides an objective of continuity of data from ERS and Envisat missions.

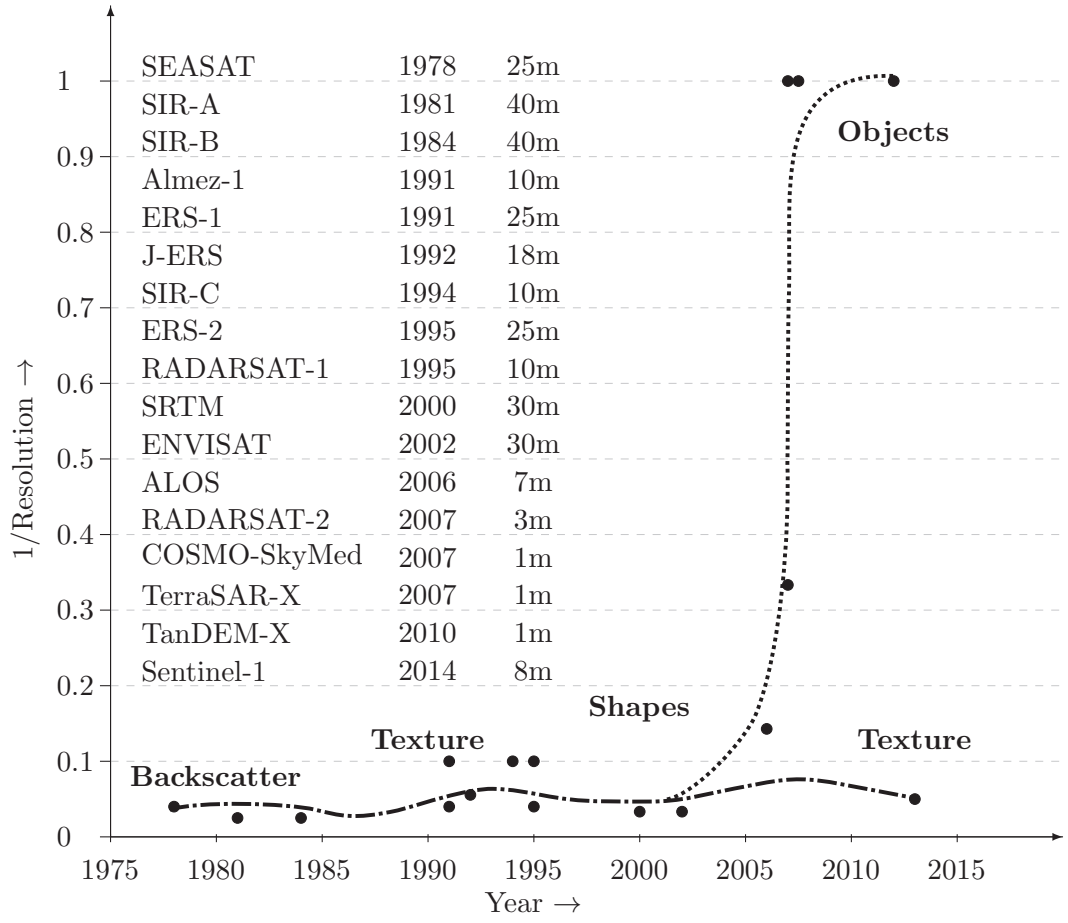


Figure 3.1: Ever increasing resolution from SAR satellites with time. The increase in resolution is changing the problem statement in SAR image understanding. The interest in metric-resolution SAR images is for detecting objects rather than land-cover classification in contrast to the low-resolution SAR images. The planned launch of Sentinel-1 in the near future with moderate resolution SAR images will not let us ignore the methods developed for such kind of images in the past.

Low-resolution images ($\sim$ 40m) are appropriate for large scale mapping and surveillance. What we obtain is the backscatter information, such as in the study of sea ice in oceanography. In a slightly better resolution, but still low ($\sim$ 20m), what seems to be visible is the textural information. Different targets, or in other words, different regions seem to look different in terms of spatial variability in neighboring pixels. In medium resolution images ($\sim 3-10$m), the shapes of the targets start to be more prominent. The recognition of smaller objects, however, still remains challenging. With high-resolution and very-high-resolution satellite images ($\sim 0.5-3$m), it is even possible to do an individual target analysis in more details. The targets of interest are then not just regions, but in several cases are objects. But this enhancement in resolution comes at the cost of huge reduction in swath of full-scene. This makes large scale reconnaissance difficult and rather specific geographical regions are emphasized.

TerraSAR-X, Acquisition: 11-OCT-2008

ERS-1, Acquisition: 24-JUL-1992

Figure 3.2: Example of the tremendous increase in details with growing resolution. ERS image of a part of the city of Bucharest (Romania) at a resolution of 25 meter (512×512 pixels), (inset) TerraSAR-X image at a resolution of ≈ 1 meter (512×512 pixels) of the area marked by the white-rectangle in (left). Both of the images are shown at the same scale in comparison to the respective resolution.

The two curves in the figure 3.1 summarize an important trend in the evolution of image resolution from the past and future SAR satellite missions. The lower curve ($- \cdot -$) which represents the low-resolution images since the launch of ERS-1 and seems to continue with the launch of Sentinel-1 in the near future is very important for large scale monitoring. Whereas the second curve ($\cdots$) signifies the quest for a better resolution in order to move from mere large scale monitoring to detailed target understanding and recognition with an increase in availability of details. Figure 3.2 shows an example of this tremendous increase in details with growing resolution. Figure 3.2 (left) is an ERS-1 image of city of Bucharest (Romania), where largely, only textural information was available. It is possible to make out the regions based on urban area, river, grassland and so on. In the inset of the figure 3.2 is shown a very-high-resolution image from TerraSAR-X, with the illustration of the scatterers over the building (Parliament house, Bucharest (Romania)) in much greater details.

3.1.1 SAR image content

3.1.1.1 Low-resolution and medium resolution SAR images

One of the biggest advantage of low-resolution SAR images is the huge area coverage due to a large swath at the cost of reduced resolution. This leads to the possibility of land applications, like soil moisture and vegetation monitoring and oceanographic applications. These were among the earliest applications of SAR images since the launch of Seasat and the space shuttle mission. In the land applications, the intrinsic spatial variability, or *texture*, of the image, has been used as a valuable feature in discriminating among different land-use types (Ulaby *et al.* [1986]). Different land-use categories tend to exhibit different texture due to different backscattering properties and varied backscattering phenomena occurring in a resolution cell as explained in the section 2.1.2.1. Hence the texture originating mainly depends upon the scene under consideration. Another important application area is the monitoring of ocean surface patterns. An early study on mapping of oceanographic structural and topographic features and observation of ocean surface waves, internal waves, current boundaries from the Seasat images have been summarized in Elachi [1980].

SAR images, like any other coherent imaging systems, are affected by a noise-like granular aspect, known as *speckle*. The understanding of fully-developed speckle where the assumption of a large number of scatterers in a resolution cell to conform with the central limit theorem, leads to many statistical models for the analysis of texture in SAR images as presented in the figure 2.3.

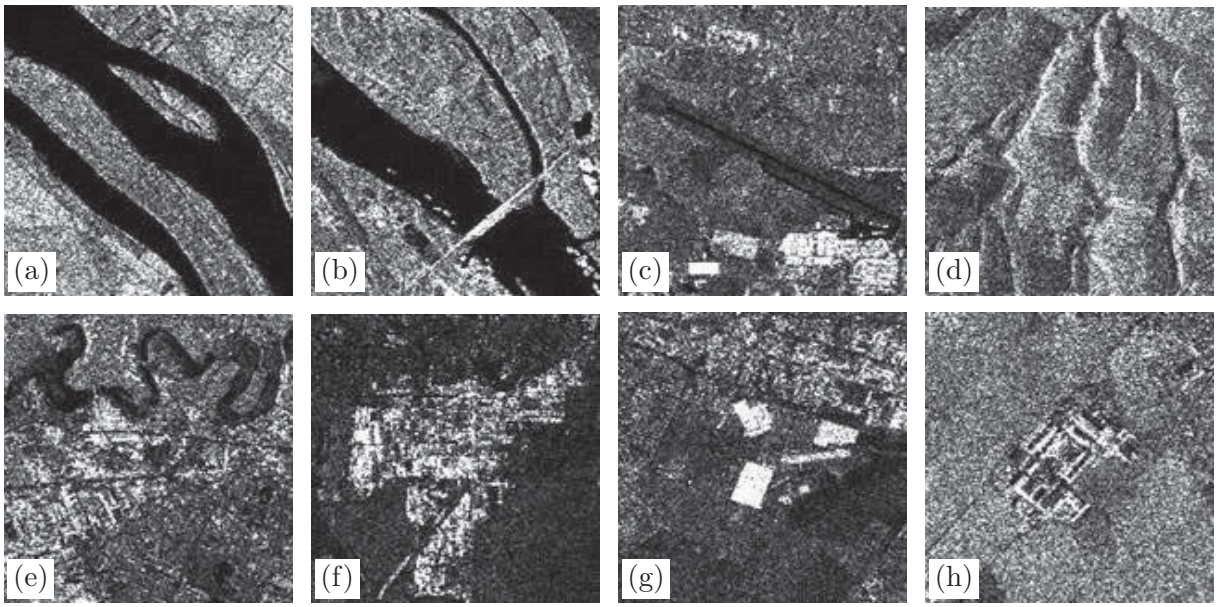


Figure 3.3: Examples from low-resolution satellite SAR images, pixel size is 256 × 256.

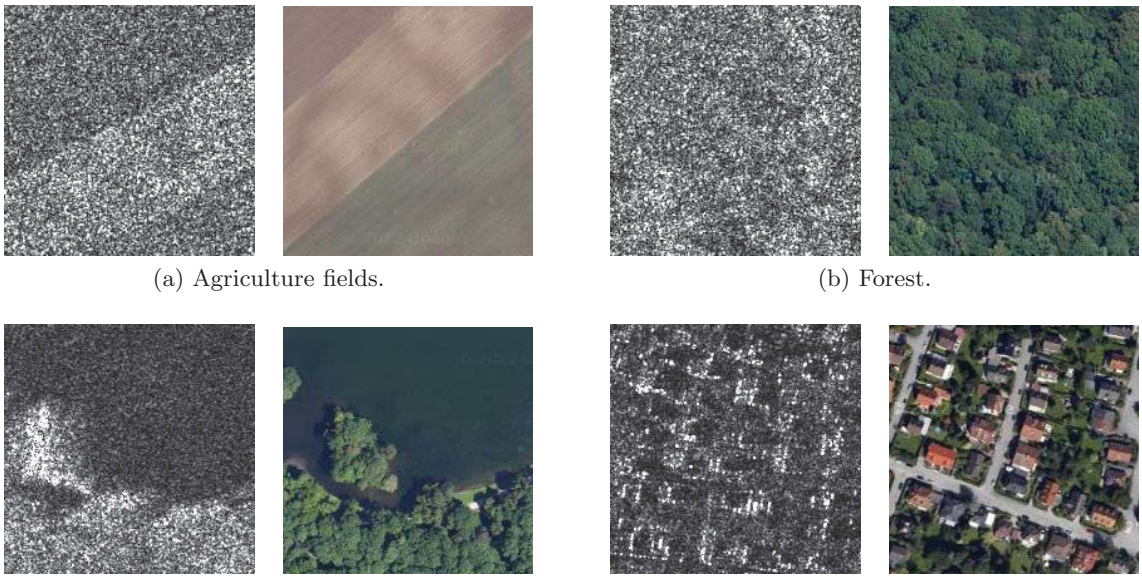
In the figure 3.3 are presented some of the typical examples of texture visible in a low-resolution SAR image. Each SAR image is of 256 × 256 pixels size extracted from a full-scene of ERS-1 satellite image (by multilooking from type SAR_IMS_1P). Examples of rivers and vegetation are presented in (a) and (b). As can be seen, rivers look very dark due to specular reflection. Considering the example of (b), along with the river and vegetation, the other interesting object is a bridge over the river, which looks exceptionally bright due to the strong backscatter from the sharp-edged concrete structure. Airstrip in the middle along with urban

set-up in the bottom left is presented in (c). Again, the airstrip also looks dark due to specular reflection from the smooth concrete layer on ground. In (d), layover and shadow effects over a mountainous region are visible. We can draw similar observation about bright looking urban set-up surrounded with land-cover typologies, both describing a specific texture in (e)-(g).

It must also be noted that apart from texture, it is also possible to carry out individual target analysis, provided that the target is large enough as compared to the resolution cell and relatively isolated from adjoining targets. For example the bridge in (b) and airstrip in (c) are the objects of concern. These objects are surrounded with well textured vegetation and the river in (b); and vegetation and urban area in (c). Similar observation can be made on an isolated building surrounded by grass fields in (h).

3.1.1.2 High-resolution and very-high-resolution SAR images

Considering the cases of high-resolution and very-high-resolution SAR images, it is still possible to detect a texture, specifically in a homogeneous region such as a uniform forest, grassland and sandy plains. Here speckle can be still considered to be fully developed. Speckle tends to be fully developed when the number of scatterers in a resolution cell is large and size of scatterers is smaller as compared to the scale of the wavelength of the incident radar wave. In this situation, when the size of scatterers / objects are much smaller than the wavelength of the incident radar wave, many scatterers will be present in a resolution cell, and thus speckle will be prominently visible in the image, degrading the image's understanding possibilities. Statistical characterization of speckle can be carried out by knowing the fact that speckle is nothing but the interference between the coherent returns from numerous scatterers present in a resolution cell. More importantly, it is this speckle phenomena which gives rise to a specific image texture. Few such examples from very-high-resolution SAR are presented in the figure 3.4.



(a) Agriculture fields.

(b) Forest.

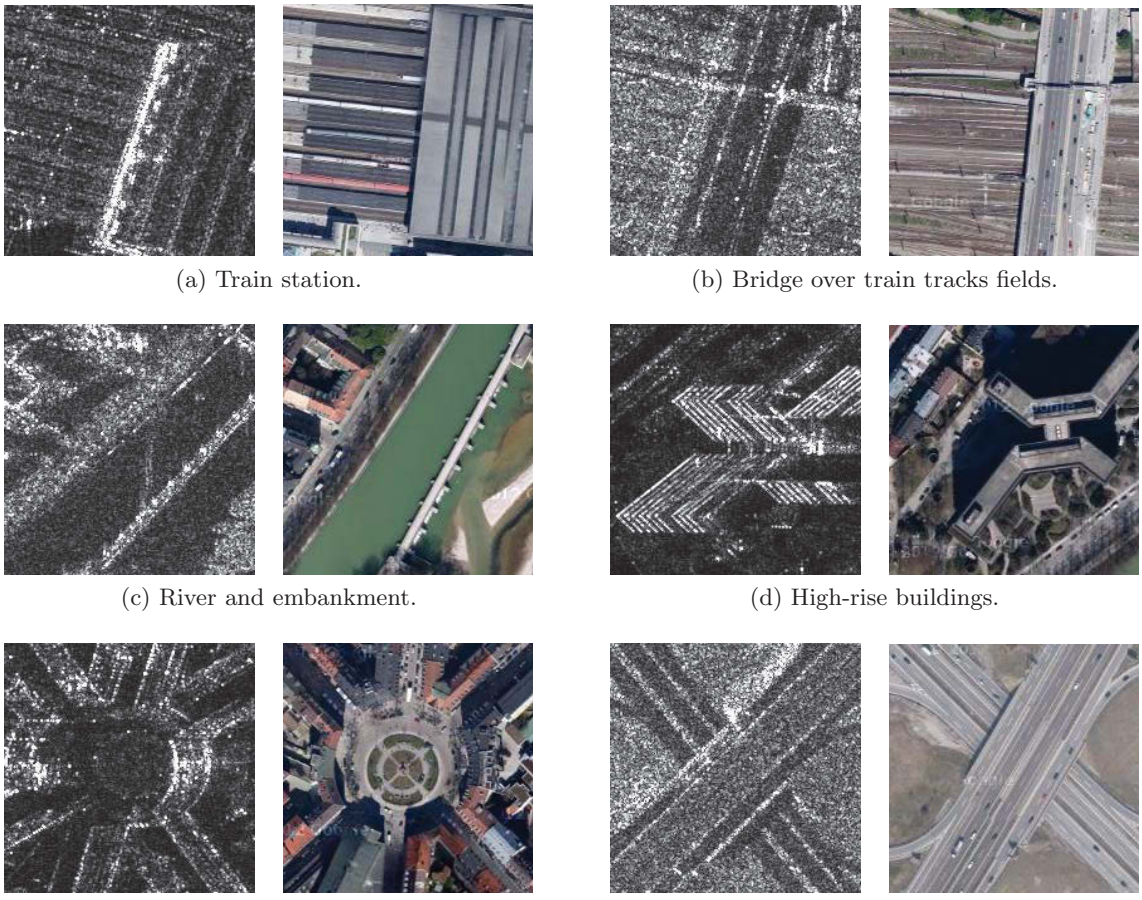
(c) Water body and vegetation.

(d) Densely populated urban setup.

Figure 3.4: Examples from high-resolution SAR images (left) in various categories with well behaving texture along with their corresponding optical satellite images (right). SAR image type: very-high-resolution spotlight mode of TerraSAR-X. Optical satellite images © Google Inc.. Geographical location-Munich(Germany).

Figures 3.4a and 3.4b present homogeneous textured regions from an agriculture field and forest, respectively. Two different types of textures can be noticed in the figure 3.4a, mainly due to bare-soil from the top-left corner and crop in the bottom right corner. Figure 3.4c shows a forest region along a water body. The water surface also looks very dark like is low-resolution SAR images, but as the size of resolution cell is comparatively larger in high-resolution SAR, the roughness of water is visible. Figure 3.4d shows a densely populated urban area surrounded with trees. On a larger scale, this also forms a kind of texture with vegetation cover embedding the bright structure due to stronger backscatter from the buildings.

The situation is largely different when we consider the regions which contains objects of relatively larger dimensions, as shown in the figure 3.5.



(a) Train station.

(b) Bridge over train tracks fields.

(c) River and embankment.

(d) High-rise buildings.

(e) Traffic roundabout.

(f) Freeway crossings.

Figure 3.5: Examples from very-high-resolution SAR images (left) in various categories with well defined objects along with their corresponding optical satellite images (right). (SAR image type: very-high-resolution spotlight mode of TerraSAR-X, Optical satellite images © Google Inc., Geographical location-Munich(Germany)).

None of the very-high-resolution SAR images presented in the figure 3.5 exhibit a specific texture on a broader scale, whereas they reveal even more detailed information about the scatterers on the ground. In the figure 3.5a as well as in the figure 3.5b, it is even possible to count the number of train tracks at the train station. A closer inspection of the figure 3.5c provides detailed information about the gates in the barrage of the embankment of a river. The strong reflection from the multiple floors above ground in a high-rise building in the figure 3.5d can

be used to determine the height of the building, provided the satellite sensor parameters are available. The distinction between a traffic roundabout in the figure 3.5e and a free-way crossing in the figure 3.5f is evident in the very-high-resolution SAR images.

With these examples, the need to reconsider the concept of speckle in such scenarios where the size of scatterers in a resolution cell is relatively very large, must be questioned. It will not be an overstatement if we say that there is no speckle in the images, whereas the observed pixels are actually the true backscatter from individual scatterers through single, double or multiple bounces. And even in the case of multiple bounces, the interference between various backscatters will not be a nuisance parameter, but valuable information. So, it is time to redirect our focus from the texture synthesis on a larger scale to individual target analysis.

Special examples in high-resolution and very-high-resolution SAR In the earlier examples, we examined some typical texture and objects in high-resolution and very-high-resolution SAR images. Now we will consider the case of some typical phenomena in SAR images occurring because of the signal processing techniques applied in SAR synthesis to achieve a higher resolution. These phenomena, specifically Doppler effects from moving targets and azimuth ambiguities, were also present in low-resolution SAR, but become evidently more prominent in the case of high-resolution SAR images. These phenomena lead to the formation of certain signatures in the SAR image on the location where there was actually no structure present on ground. To understand this we need to look into the complex-valued SAR images, i.e. into real and imaginary components of SAR images.

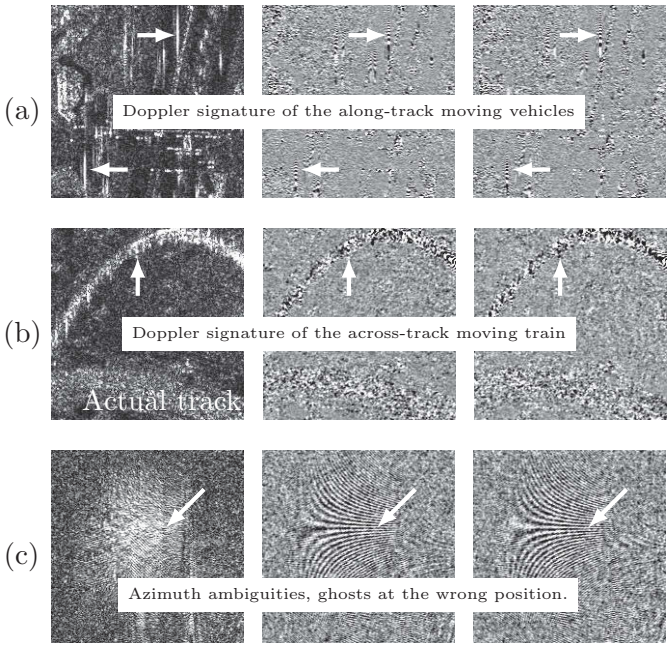


Figure 3.6: (Left) Amplitude, (center) real component, and (right) imaginary component from the Doppler signature from the moving targets in (a) and (b), and azimuth ambiguities over a sea-surface from high rise building at adjoining coastal area in (c), marked with a white arrow. It can be seen, its not possible to discriminate these signatures from other targets in the amplitude image, so we should consider analyzing them in complex-valued space by utilizing the embedded phase information.

The Doppler signature of the target moving along the track appears smeared in azimuth (figure 3.6a), whereas the Doppler signature of the targets moving across track appears displaced

in azimuth from the actual target location (figure 3.6b), (Hinz *et al.* [2007]). Doppler chirp can be seen in the real and imaginary components in both cases. Azimuth ambiguities in SAR are caused by the aliasing of the Doppler phase history of each target that is sampled according to the sensor azimuth sampling frequency, which equals the pulse repetition frequency (PRF) (Guarnieri [2005]), briefly introduced in section 2.2.2. The ghosts at the wrong position from the target, that are actually not present on that location on ground are shown in the figure 3.6c. A specific signature in real and imaginary components will make the analysis of such ambiguities possible. It must be noted that though not in all, but at least in certain specific cases, the phase information in the complex-valued SAR images is of significant importance.

3.2 Individual target understanding in SAR images

We noticed a variety of examples of targets (textures and objects) in sections 3.1.1.1 and 3.1.1.2 for low-resolution and high-resolution SAR, respectively. We also commented that it is possible to carry out a more detailed individual target analysis in case of high-resolution SAR images, as the information available for each target is sufficient enough for individual target analysis. Nevertheless, methods of automatic information extraction and objects understanding are always in demand, a large section in application area rely upon the visual interpretation of the targets for a first hand analysis. In this section, we propose a visual exploration method (Singh & Datcu [2012c]), with an underlying idea based on a kind of *hyper-image* (Bertrand & Bertrand [1996]) generation in a 4-D space. This space represents the spatial coordinates (x, y) and the frequency coordinates (f_r, f_a), in range and azimuth, respectively. The proposed method aims to assist image analysts in differentiating various SAR targets by *looking* into the details of the spectrum, which might have different scattering behavior, but otherwise visually look similar in spatial domain.

This method will not only provide an interpretation tool to SAR image analysts, but will also help us understand the intrinsic properties of images with SAR systems with larger bandwidth (such as TerraSAR-X) and forms a basis for further conclusion drawn in this thesis on the SAR image understanding. During SAR imaging, in low-resolution SAR, response of targets can be assumed to remain unchanged during illumination time and transmitted chirp bandwidth. But in the high-resolution SAR, the backscattering from various targets varies in azimuth chirp bandwidth and this variation is embedded in the raw data. This variation can be used as the signature for various targets. Complex-valued SAR images, obtained after focusing of raw SAR data also carry this information. As the availability and processing of raw SAR data is not very convenient, we have considered and studied as methods to extract this information from focused SAR images.

3.2.1 Theoretical background

We discussed the concept of azimuth splitting in section 2.2.2. Azimuth splitting (or azimuth sub-band/sub-aperture in the complex spectrum of the SAR signal) corresponds to the splitting of the full synthetic aperture in order to extract a specific viewing angle of the sensor. Each Doppler phase term in the Doppler effect generated due to the radar/target range variations during the SAR acquisition is related to a specific sensor position and hence to a specific azimuth viewing angle. This concept is based on the principle exposed in Arnaud [1999] for ship detection, using interferometric degree of coherence between two azimuth sub-looks of SLC, namely front look and rear look. In Schneider *et al.* [2006], a method of detection of coherent scatterers (CSs) is proposed using sub-look coherence (SC) and sub-look entropy (SE). Spectral correlation

is carried out only for azimuth sub-looks. Later SC and SE is compared with interferometric coherence and polarimetry entropy respectively for characterization of CSs. In Kim *et al.* [2003], a sub-look pyramid is defined, in order to model framework incorporating scatterer's anisotropy. In Tison & Fjortoft [2008], and Spigai *et al.* [2011], a continuous sub-band analysis is proposed, which relies on splitting the full-aperture into continuous sub-looks both in range and azimuth directions.

The concept of azimuth splitting method has been extended to range direction in a number of studies. In Souyris *et al.* [2003], the two-look internal Hermitian product (2L-IHP) is defined from a series of sub-looks generated from SLC image, in order to study the inherent speckle effects. The study presented in this section is an extension of the principles of continuous sub-look decomposition proposed in Tison & Fjortoft [2008], and Spigai *et al.* [2011]. It involves simple band-pass filtering of a spectrum with a continuous displacement of the analyzing window to obtain a 4-D array, as shown in the figure 3.7.

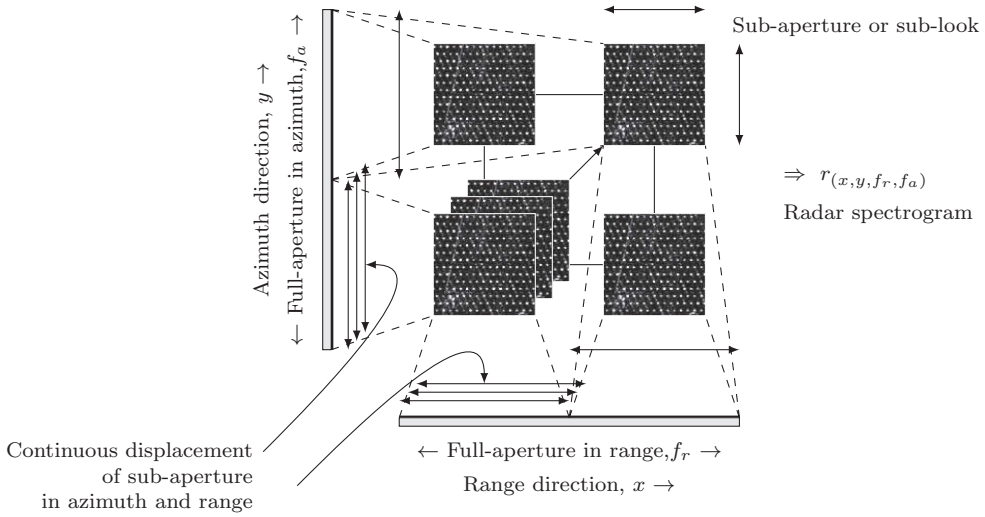


Figure 3.7: Multiple sublook decomposition. The extension of azimuth splitting to the range direction in order to generate a 4-D matrix $r_{(x,y,f_r,f_a)}$, denoted as *radar spectrogram* in Spigai *et al.* [2011] .

This 4-D array is then visualized in 2-D. This visualization is similar to hyper-images proposed in Colin *et al.* [2004]. Simple band-pass filtering preserves a direct analogy between azimuth/range frequencies and antenna sub-look/chirp bandwidth (Tison & Fjortoft [2008]), and a 2-D representation provides a physical interpretation of the scattering behavior of the target. Some important observations on the representation of a 4-D array in 2-D for high-resolution TerraSAR-X are presented in Singh & Datcu [2010], leading to the proposed method of target analysis based on a 4-D representation by means of slicing and animations for visual inspection of the scattering behavior of the target under consideration.

3.2.2 Visual exploration based on 4-D representation of multiple sub-look decomposition (MSLD)

Visual inspection of the targets in SLC images may provide incomplete and/or misleading information, as sometimes two targets having entirely different scattering behavior on ground, might look similar in a SAR image. In this section, we provide the concept behind the proposed method, which aims at providing additional information to image analysts, derived using sub-look decompositions of a SLC image, for discrimination of targets. Realization of this 4-D

representation method is based on studies carried out in Tison & Fjortoft [2008], Spigai *et al.* [2011], and Singh & Datcu [2010]. In Tison & Fjortoft [2008], and Spigai *et al.* [2011] proposed that TFA performs continuous sub-look analysis to generate a 4-D array $r_{(x,y,f_r,f_a)}$, termed as *radar spectrogram*, which can be written as follows:

$$r_{(x,y,f_r,f_a)} = \mathrm{FT}^{-1}\left[w_{(f_r,f_a)} \times \mathrm{FT}(S)\right](x,y). \tag{3.1}$$

Here, (x, y) refer to spatial positions in range and azimuth, and (f_r, f_a) refer to range and Doppler frequencies. Band-pass filter $w_{(f_r,f_a)}$ is a Hamming filter cantered at (f_r, f_a) with bandwidths bw_r in range and bw_a in azimuth. S is the complex-valued SAR image of size N × N pixels containing the target of interest and contextual information. FT is the Fourier transform and FT^{-1} the inverse Fourier transform. The spectrum of S, i.e. $\mathrm{FT}(S)$ has full aperture bandwidths BW_r in range and BW_a in azimuth. This process is a simple band-pass filtering of the complex-spectrum in range and azimuth direction. In Tison & Fjortoft [2008], and Spigai *et al.* [2011] *radar spectrogram* is computed for every pixel in the scene. A 2-D image of *radar spectrogram* corresponding to the center pixel, i.e. $r_{(\mathbf{x},\mathbf{y},f_r,f_a)}$ with $\mathbf{x} = \mathrm{N}/2$ and $\mathbf{y} = \mathrm{N}/2$ is used to generate a global classification. In our study we are mainly interested in detailed analysis of particular targets by displaying $\hat{r}_{(x,y,f_r,f_a)}$ in the form of animations and slices, where $\hat{r}_{(x,y,f_r,f_a)} = \mid r_{(x,y,f_r,f_a)} \mid$.

We start with an SLC image. Doppler centroid is estimated (Madsen [1989]) and corrected as a preprocessing step. As this is a visual exploration method, accuracy of Doppler centroid estimation is not very critical. After Doppler centroid correction, $\hat{r}_{(x,y,f_r,f_a)}$ is computed for the selected area of interest using the equation 3.1. Computed $\hat{r}_{(x,y,f_r,f_a)}$ is stored in a 4-D matrix. A 2-D image of data in the 4-D matrix can be created by the slice of any two dimensions while keeping the other two dimensions fixed. This 2-D image can be animated by changing one of the fixed parameters enabling visualization of three dimensions of the matrix. There are 12 possible animations / slices, but only 6 are non-repetitive.

Table 3.1 summarizes the description of these possible slices. The abscissa and the ordinate represent different parameters in every slice. We take TerraSAR-X high-resolution spotlight mode data with the 300 MHz option for our experiments. The range chirp has a band-width (BW_r) of 300 MHz. The carrier frequency of 9.65 GHz is modulated in azimuth by approximately 3 kHz (Mittermayer & Runge [2003], CAF-DLR [2009]). So azimuth chirp has a band-width (BW_a) of 3 kHz. We also assume that the time-variant Doppler spectrum ramp (Eineder *et al.* [2009]) has already been corrected. The MSLD is carried out using Hamming filter with band-widths $bw_r = BW_r/2 = 150$ MHz and $bw_a = BW_a/2 = 1.5$ kHz approx.. Selecting $N = 128$ pixels provide us a 4-D matrix of dimensions $65 \times 65 \times 65 \times 65$ pixels. Values for N, bw_r and bw_a can be user specified, but the interpretation of results also needs to be adopted accordingly.

Slice$_1$ $\left(\hat{r}_{(x,y,\mathbf{f_r},\mathbf{f_a})}\right)$ is the low-resolution image seen from a particular sub-look $(\mathbf{f_r}, \mathbf{f_a})$, analogous to range chirp and azimuth chirp. The animation will show all the low-resolution images seen from all the sub-looks one after the other, thus providing an insight to how a particular analyzing area is visible to different looks (range and azimuth) of the synthetic antenna. Highly directive targets will not be visible in all frames of animation (i.e. in all sub-looks), whereas coherent or isotropic targets will be visible in all the frames. The abscissa and the ordinate represent the range and the azimuth respectively (both in meters). MSLD with a filter of band-width half of the full-aperture band-width results in reduction of the resolution by half. With original resolution of approx. 1 m, the abscissa represents $x = 0, 2, \ldots 128$ and the ordinate represents $y = 0, 2, \ldots 128$.

Slice$_2$ $\left(\hat{r}_{(\mathbf{x},\mathbf{y},f_r,f_a)}\right)$ shows how a particular point $(\mathbf{x}, \mathbf{y})$ on the ground is seen from all sub-looks. The band-pass filter centered at (f_r, f_a) with band-widths $bw_r = BW_r/2$ and $bw_a =$

Table 3.1: Description of various slices

Slice_1	$\hat{r}_{(x,y,\mathbf{f_r},\mathbf{f_a})}$, Sub-look image
	(Low-resolution image seen from a particular sub-look)
Slice_2	$\hat{r}_{(\mathbf{x},\mathbf{y},f_r,f_a)}$, Radar spectrogram
	(Particular point on ground seen from all sub-looks)
Slice_3	$\hat{r}_{(x,\mathbf{y},\mathbf{f_r},f_a)}$, Range line vs. Doppler frequency
	(Behavior of a particular line at fixed azimuth seen from different azimuth sub-looks)
Slice_4	$\hat{r}_{(x,\mathbf{y},f_r,\mathbf{f_a})}$, Range line vs. Range frequency
	(Behavior of a particular line at fixed azimuth seen from different range sub-looks)
Slice_5	$\hat{r}_{(\mathbf{x},y,f_r,\mathbf{f_a})}$, Azimuth line vs. Range frequency
	(Behavior of a particular line at fixed range seen from different range sub-looks)
Slice_6	$\hat{r}_{(\mathbf{x},y,\mathbf{f_r},f_a)}$,Azimuth line vs. Doppler frequency
	(Behavior of a particular line at fixed range seen from different azimuth sub-looks)

$BW_a/2$ is applied in range and azimuth. Representing only the center frequency on the axes, the extent of the abscissa ranges from 75 MHz to 225 MHz and the ordinate ranges from 0.75 kHz to 2.25 kHz.

Slice$_3$ $\left(\hat{r}_{(x,\mathbf{y},\mathbf{f_r},f_a)}\right)$ shows the behavior of a particular line $\mathbf{y}$ of fixed azimuth in spatial domain, seen from different azimuth sub-looks but at a fixed range sub-look, $\mathbf{f_r}$. The range on the abscissa extend to 128 m and the center frequency of the band-pass filter applied to the Doppler chirp on the ordinate ranges from 0.75 kHz to 2.25 kHz.

Slice$_4$ $\left(\hat{r}_{(x,\mathbf{y},f_r,\mathbf{f_a})}\right)$ shows the behavior of a particular line $\mathbf{y}$ of fixed azimuth in spatial domain, seen from different range sub-looks but a fixed azimuth sub-look, $\mathbf{f_a}$. The range on the abscissa extend to 128 m and the center frequency of the band-pass filter applied to the range chirp on the ordinate ranges from 75 MHz to 225 MHz.

Slice$_5$ $\left(\hat{r}_{(\mathbf{x},y,f_r,\mathbf{f_a})}\right)$ shows the behavior of a particular line $\mathbf{x}$ of fixed range in spatial domain, seen from different range sub-looks but a fixed azimuth sub-look $\mathbf{f_a}$. The azimuth on the abscissa extend to 128 m and the center frequency of the band-pass filter applied to the range chirp on the ordinate ranges from 75 MHz to 225 MHz.

Slice$_6$ $\left(\hat{r}_{(\mathbf{x},y,\mathbf{f_r},f_a)}\right)$ shows the behavior of a particular line $\mathbf{x}$ of fixed range in spatial domain, seen from different azimuth sub-looks but a fixed range sub-look $\mathbf{f_r}$. The azimuth on the abscissa extend to 128 m and the center frequency of the band-pass filter applied to the Doppler chirp on the ordinate ranges from 0.75 kHz to 2.25 kHz. Thus, if a target is varying in time on a particular line of fixed range, then it will be seen differently by different parts in azimuth chirp, in turn generating a special signature in this slice.

Slice_1 to Slice_6 can be used in the form of images and/or animations to differentiate different targets, presented in the next section.

3.2.3 Illustrations of examples

In this section we suggest methodologies to interpret and analyze targets by manual inspection. However, because of high complexity of data and diversity of targets in the high-resolution SAR, the user must learn about targets of interest *a priori* in order to apply this knowledge to interpret and recognize unknown targets. We will demonstrate a few of the various possible examples. We propose three methods to use $\hat{r}_{(x,y,f_r,f_a)}$ for differentiating the targets. The first

method is an animation visualization tool, the second is the *radar spectrogram* method. The third method uses the slices of $\hat{r}_{(x,y,f_r,f_a)}$ on different possible planes, i.e. using all the possible 2-D representations of $\hat{r}_{(x,y,f_r,f_a)}$. These three methods can be used in synergy or independently, depending on the target of interest.

Visual interpretation using animated Slice$_1$

Slice$_1$ ($\hat{r}_{(x,y,\mathbf{f_r},\mathbf{f_a})}$) is the low-resolution image seen from a particular sub-look ($\mathbf{f_r}, \mathbf{f_a}$), analogous to range chirp and azimuth chirp. The animation shows all the low-resolution images as seen from all the sub-looks one after the other, providing an insight into how a particular analyzing area is visible at different looks (range and azimuth) of the synthetic antenna at the cost of resolution. Coherent targets are visible in all the frames of the animation.

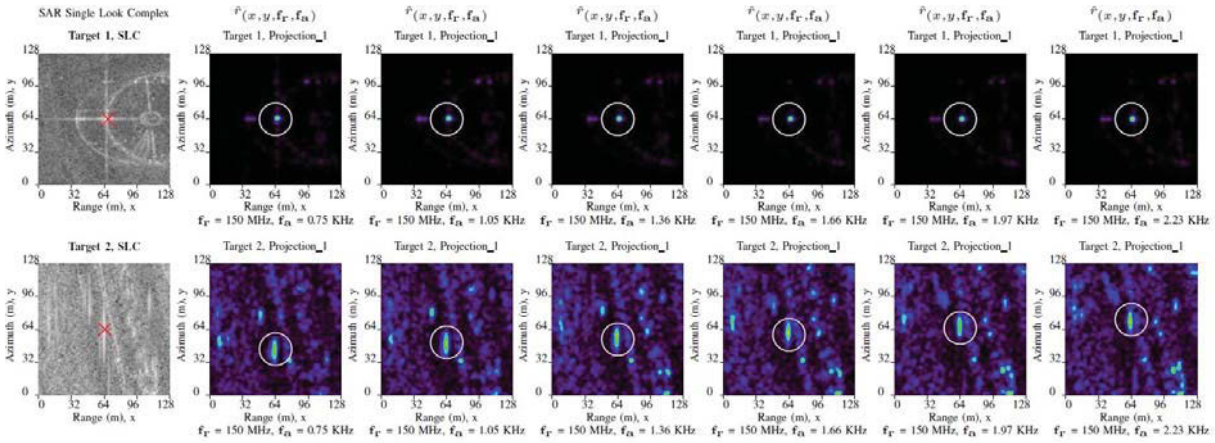


Figure 3.8: Visual inspection using animation of Slice$_1$ ($\hat{r}_{(x,y,\mathbf{f_r},\mathbf{f_a})}$). Here, (x, y) refers to spatial positions in range and azimuth and $(\mathbf{f_r}, \mathbf{f_a})$ refers to fixed range and Doppler frequencies. Randomly selected 6 frames from Slice$_1$ ($\hat{r}_{(x,y,\mathbf{f_r},\mathbf{f_a})}$) for two targets are presented. First row: **Target-1**, marked by red crosses, consisting of masts at oil-containers. The animation shows no spatial variation in signatures (enclosed by white-circles in first row) of stable targets for different range and azimuth sub-looks. Second row: **Target-2**, marked by a red cross, for a region with vertical wall like structures on a highway. Signatures of targets can be seen moving in up-wards direction when seen in animation of Slice$_1$ for different range and azimuth sub-looks (see the motion of signatures enclosed by white-circles in second row), thus this target can be assigned as signature of Doppler shift due to a vehicle moving in Azimuth Direction.

In the first example, we analyze two targets shown in the figure 3.8. The bright spots in the **Target-1** are masts on the oil-containers, assumed to be coherent targets. Six arbitrarily selected frames of the animation obtained from Slice$_1$ are presented in the figure 3.8. The point targets are visible in all the frames and at the same spatial location. This implies that the backscatter from these target is seen from all the sub-look of antenna at the same spatial location. This confirms the assumption of assigning them as coherent targets. **Target-2** is a highway where vertical structures are seen. These are difficult to be discriminated as actual structures on ground or ambiguities from moving vehicles. Animation in Slice$_1$ shows spatial variations of the target response in case of ambiguity. As can be seen from the 6 arbitrarily selected frames, in case of **Target-2**, signature moves spatially, while the stable targets remain spatially stationary. This can support differentiation of actual structure and moving targets.

Visual interpretation using Slice_2

This slice ($\hat{r}_{(\mathbf{x},\mathbf{y},f_r,f_a)}$) is similar to the *radar spectrogram* (Tison & Fjortoft [2008], Spigai *et al.* [2011]) computed for a particular pixel $(\mathbf{x},\mathbf{y})$. Slice_2 here includes the effects of contextual information because of larger analyzing window size (aw). However, aw should not be too large to pollute the punctual target signal. It can be very effective in analyzing the isolated targets.

Visual interpretation using Slice_1 to Slice_6

Although Slice_1 and Slice_2 provide information about the scattering variation of targets, in order to get detailed physical interpretation, it is useful to go into the details of all the other slices too. In the preceding section we have provided the physical significance of all the slices summarized in the table 3.1. In some targets, the effects visible in Slice_1 animation can be interpreted directly from one particular frame of the other slices, but these need to be interpreted differently as explained later in this section. In the figure 3.9, various targets are compared using Slice_1 to Slice_6. Targets can be seen marked by a red cross.

Target-3 and **Target-**4 (first and second row of the figure 3.9) are taken from the same parking lot and look similar in the SLC image. Ground truth using Google-Maps® demonstrates an electrical pole existence at the center of **Target-3**, while no such target is present in **Target-**4. In the animation of Slice_1 for **Target-3**, the signature remains spatially stationary, while it is variable for **Target-**4. Slice_2 for **Target-3** represents the unmodified antenna pattern while response of **Target-**4 modifies this pattern. Slice_3 to Slice_6 for **Target-3** show uniform variation, while slanted signature in Slice_6 for **Target-**4 represents variation in response for a fixed range but different azimuth sub-looks. These observations confirm the assumptions that **Target-3** is a meaningful structure, but **Target-**4 is ambiguous. **Target-5** and **Target-6** (third and fourth row of the figure 3.9) are taken from two different highways, containing a wall like structure along the azimuth direction, which could be a traffic separators at highway in both the case. Analyzing all the slices in both cases, we find a unique slanted signature in Slice_6 for **Target-5**, with such phenomenon missing in Slice_6 of **Target-6**. This slanted signature shows that the response of the target is highly variable for the fixed range when seen from different azimuth sub-looks. So we can confirm that the vertical structure in **Target-5** is Doppler shift due to a target moving in azimuth direction, and the vertical structure in **Target-6** is a wall like structure giving a strong backscatter. **Target-7** (fifth row of the figure 3.9) is also taken from a highway. Two bright points in SLC image of this area can be seen, but ground truth from Google-Maps® confirms that there exists no such structure at that point of the highway. Just by looking at the SAR image, it is not possible to find out the exact nature of these two targets. But two slanted lines in Slice_6 for **Target-7** confirm that these targets are vehicles moving in opposite direction in azimuth. The same is confirmed by looking at the animation of Slice_1 where two broad points moving in opposite directions in different frames (the upper one downwards and the lower one upwards) can be seen. **Target-8** (last row of the figure 3.9) is taken from sea area in order to better explain the strong ambiguities due to a side-lobe, as its difficult to locate ambiguities in other regions. Typically a slanted signature in Slice_3 and Slice_5, and a blob like structure moving in horizontal direction in the animation of Slice_4 and Slice_6 exists in case of such ambiguities.

3.2.4 Conclusion

The presented examples are special cases, but MSLD can be applied to a large set of targets and objects in similar manner. The objective behind discussing this visual exploration approach

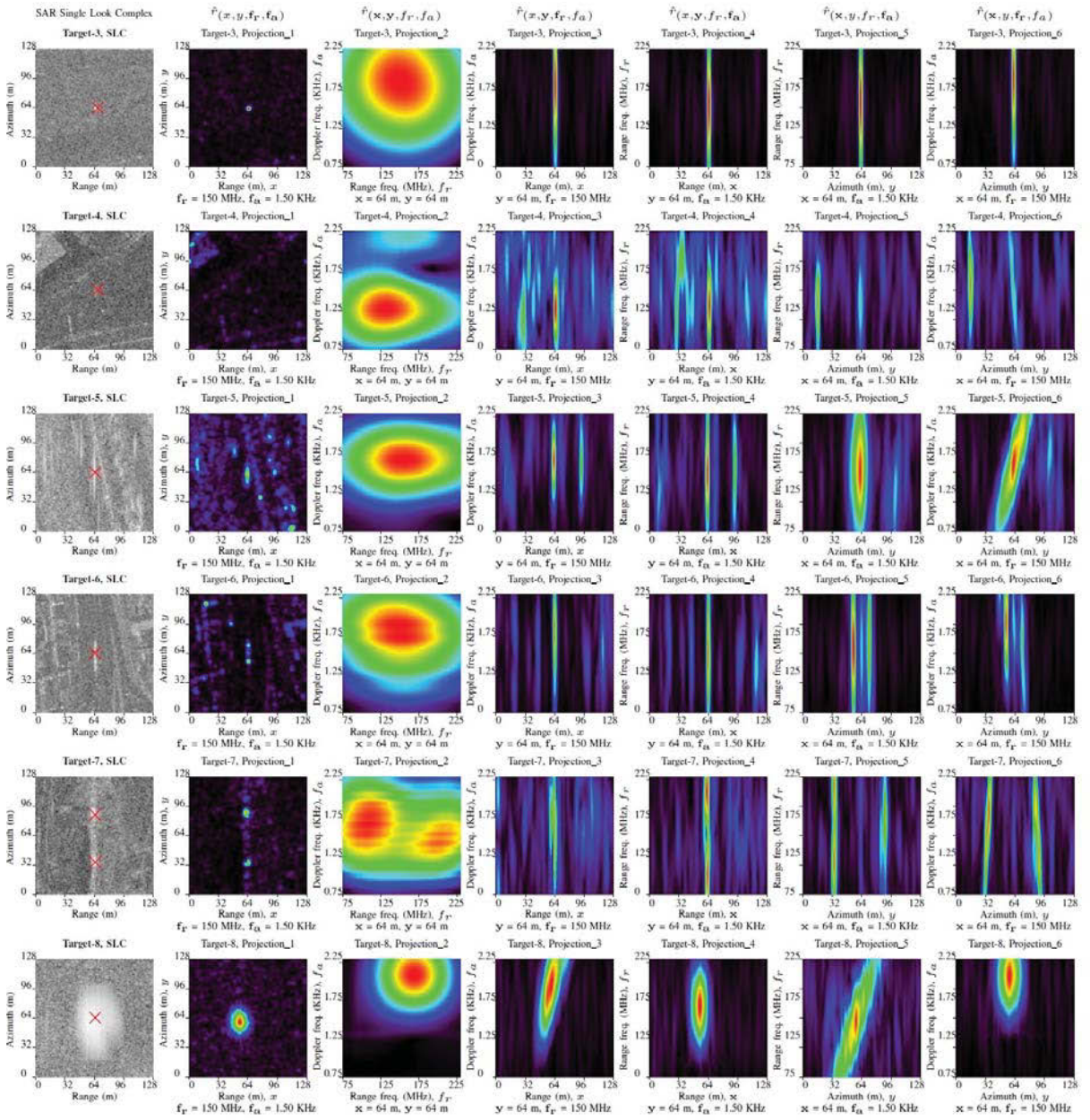


Figure 3.9: Visual inspection using all the slices of 4-D array $\hat{r}_{(x,y,f_r,f_a)}$. Slices are generated by visualizing two dimensions on a plane keeping other two dimensions fixed. In above figure, (x, y) refers to spatial positions in range and azimuth, (f_r, f_a) refers to range and Doppler frequencies, $(\mathbf{x}, \mathbf{y})$ refers to fixed spatial positions in range and azimuth and $(\mathbf{f_r}, \mathbf{f_a})$ refers to fixed range and Doppler frequencies. Targets can be seen marked by a red crosses. **Target-3** and **Target-**4 are similar looking target on a same parking area with corresponding slices in first and second row. Slice_2 for **Target-3** is similar to antenna-pattern because of uniform response from target at center in all range and azimuth sub-looks. Slanted signature in Slice_6 of **Target-**4 represents the variation of response of target with azimuth sub-look for a line of fixed range. **Target-5** and **Target-6** are similar looking target on highways with corresponding slices in third and fourth row. Presence of slanted signature in Slice_6 of **Target-5** confirm that its an indication of Doppler shift due to a target moving in Azimuth direction, however missing of this phenomenon in Slice_6 of **Target-6** helps us assign it as an actual structure present on ground. **Target-7** is another target on a Highway with corresponding slices in the last row. Spatially varying bright spots in animation of Slice_1 for this target as well as slanted signature in Slice_6 help us to assign the bright spots in SLC as moving vehicles on highway moving in opposite directions. **Target-8** can be assigned as an ambiguity due to side-lobes of SAR antenna by visualizing the slanted signature in Slice_3 and Slice_5. Ambiguities due to side-lobes also exhibits a specific moving blob-like signature in the animation of Slice_4 and Slice_6.

was to generate a motivation for the use of phase information in complex-valued SAR images, which is otherwise ignored in many SAR applications, except for interferometric and polarimetric purposes. The scattering behavior of targets, which is achieved by exploiting magnitude and phase information of SLC jointly by sub-look decomposition has been addressed. So apart from statistical analysis methods primarily based on the amplitude of SAR images, we need dwell into the methods based on this spectroscopic like property of SAR.

3.3 Spatial content understanding in SAR images

We just discussed the huge diversity of images based on its content available in SAR images. Such a huge diversity emphasize the need of information extraction to describe the content of image in the form of a very compact intrinsic **feature descriptor**. This multidimensional feature descriptor acting as a signature can be later used for categorization and indexing the image content. An image is characterized by its primitive features such as texture (Tuceryan & Jain [1998]), color, and shape (Jain & Vailaya [1996]). Specifically for the case of SAR images, we may add one more important primitive feature, i.e. feature characterizing the backscattering behavior on the ground depending upon the material's dielectric properties, its orientation on ground and the size of scatterers as compared to the wavelength of the transmitted electromagnetic wave to image a surface. This property becomes even more evident with the availability of phase information in the complex-valued images, specially for the case of high and very-high-resolution SAR images where a chirp of broader bandwidth is used for imaging the surface. We already discussed some related issues in the understanding of variety of target individual target using a joint time-frequency analysis using MSLD in section 3.2. However the MSLD method is very complicated to generate a compact feature descriptor for image content. Nonetheless, this backscattering behavior will have a direct or indirect impact on the texture, color (in the case of SAR images, the intensity of gray values) and shape. So we will classify the general understanding of the spatial content to generate a **feature descriptor** for SAR images into two major aspect: time-frequency aspects and statistical aspects.

Time-frequency analysis methods

Time-frequency analysis (TFA) methods are based on general image processing paradigms. Specifically in SAR images, understanding time-frequency behavior is a direct consequence of the signal processing techniques applied to achieve better resolution, named SAR synthesis (Massonnet & Souyris [2008]). In the literature, TFA relies on three main methods: Wigner-Ville decomposition, Fourier transform or wavelet transform, by representing the images into a suitable transformation space.

The use of Wigner-Ville decomposition for the moving-target detection and imaging in SAR images was demonstrated in Barbarossa & Farina [1992]. In Popescu *et al.* [2008], the use of Wigner-Ville decomposition based on Renyi Entropy criterion was presented. But being a quadratic transform, the Wigner-Ville distribution is influenced by cross-term interference when more than one target is present (Qian & Chen [1999]).

We presented a method of spatial content understanding in SAR images based on the Fourier transform in Singh *et al.* [2009], and Popescu *et al.* [2010]. The use of a spectral feature descriptor motivated from the timbrel spectral features from music genre was developed. In Singh *et al.* [2009], the applicability of this spectral feature descriptor was demonstrated for pixel-based classification and in Popescu *et al.* [2010] as a space-variant approach for structure indexing in SAR images. Another example of the use of the Fourier transform for individual target understand-

ing has been presented in section 3.2, but this method is very complicated for its applicability to the spatial content understanding in broader sense. But based on the knowledge achieved in MSLD methods, we later proposed the method of the fractional Fourier transform (Singh & Datcu [2013]), which is actually a true scaling in phase to form the basis for a multi-scale approach in complex-valued SAR image analysis. The use of fractional Fourier transform was earlier limited largely to the moving-target detection only (Sun *et al.* [2002]). Here, the relation between the fractional Fourier transform with the Wigner-Ville distribution is derived and its advantages over the Wigner-Ville distribution demonstrated. The fractional Fourier transform is a linear transformation free from cross-term interference. We proposed to transform a SAR image in various multi-scale sub-bands with respect to a scaling in phase. Moments of these sub-bands was proposed as an obvious choice for a feature descriptor in Singh & Datcu [2012b], which was later enhanced by the use of log-cumulants as presented in Singh & Datcu [2013].

Apart from the utility of wavelets in major image compression standard (JPEG-Committee [2013]), the wavelets have been extensively used in representing the images based on color and shape. The general approach is to use the energy of the wavelet coefficients in several bands with assumption that the energy distribution in frequency domain will identify texture. Another kind of wavelet based methods such as Gabor filter banks have also been extensively covered in literature. Gabor filter banks are suitable for the analysis of signals with sharp changes but which stay for a small duration of time. The most essential property of Gabor coefficients for extracting the geometrical shape information in the image to synthesis the texture is its bandpass filter property, which has also been demonstrated in Farrokhnia & Jain [1991]. This bandpass filter like operation is also the underlying mechanism in human visual systems, which we try to imitate in the Gabor expansion of signals to understand the texture. In Bovik *et al.* [1990], as well as in Manjunath & Ma [1996], the use of Gabor filter banks to characterize the texture of images has been studied extensively. Again the moments of Gabor filtered images are obvious choice for a feature descriptor, however we proposed the use of log-cumulants of Gabor filtered images with enhanced performance over liner-moments (Singh & Datcu [2012d], Singh & Datcu [2013]).

Statistical methods

Statistical aspects pertaining to the density parameters have been introduced briefly in section 2.3. The knowledge on statistical aspects of SAR images eventually lead us to the understanding of underlying texture. We first consider a short overview on the texture synthesis, which is of interest in SAR images, at least in the homogeneous region. Later we will dwell into the applicability of more advanced Bayesian based approaches for texture synthesis, also suggested in the scope of de-speckling. The principal approaches used to describe texture has been divided into three broad categories: statistical, geometrical, and model-based in Tuceryan & Jain [1998].

The gray-level co-occurrence matrix (GLCM) can be considered as one example of statistical approach of texture characterization. The GLCM or the gray-tone spatial-dependence matrix method is a way to extract second-order statistical texture features used in various image processing applications proposed by Haralick in the 1970s (Haralick *et al.* [1973], Haralick [1979]). In Baraldi & Parmiggiani [1995], the use of the GLCM texture parameters for the texture characterization in optical satellite images (NOAA AVHRR) have been investigated. The texture features generated using the GLCM have been successfully applied to the texture analysis for mapping sea ice patterns in low-resolution ERS-1 SAR images in Soh & Tsatsoulis [1999]. The use of the GLCM in efficient texture analysis in ENVISAT-ASAR images by introducing the

notion of patch re-occurrences has been proposed in Kandaswamy *et al.* [2005]. The texture features based on the GLCM generally outperforms other classical methods of classification, at least in the case of low-resolution SAR and optical satellite images. For SAR images, the use of the GLCM is largely limited to sea-ice mapping application only.

An example of texture synthesis on the basis of geometrical approach has been presented in Tuceryan & Jain [1998]. Here, geometrical features of texture elements or primitives are analyzed by generating texture tokens based on the shape and placement rules of the Voronoi tessellation of a given image.

One of the model based approach is to first transform the image into wavelet domain in order to simplify its statistical response and then compute the density parameters from a rather simpler model, fitting the image within the transformation space (Do & Vetterli [2002]). The other methodology is to work in the image domain by model based texture synthesis. The model based texture synthesis works on the principle of modeling the scene reflectivity in the SAR image to describe the spatial structure using some complex texture models, such as Markov Random Fields (Hassner & Sklansky [1980]) and Gibbs Random Fields (Besag [1974]). An example of such a methods has been demonstrated in Datcu *et al.* [1998] by characterizing the image content using some stochastic models within a Bayesian framework. Here, the potential of using the Gibbs Random Fields models in Bayesian framework for extracting spatial textural information from remote sensing images have been focused. The applicability of Bayesian framework for statistical understanding is basically based on the multiplicative model of noise. So in order to lay a foundation of Bayesian based information extraction, it is rational to discuss some historical overview over speckle considerations.

3.3.1 Speckle and de-speckling to information extraction

Statistical understanding of SAR images is largely based on speckle characterizations. Noise-like granular aspect in SAR images, known as speckle, pose many limitations to the algorithm developments and implementations to several applications of SAR images including land use mapping, surveillance, automatic target detection and recognition etc. The presence of speckle reduces the detectability of objects in the image and also the capability to separate and classify distributed targets (Oliver & Quegan [1998]). In order to achieve a better accuracy in application results, many efforts have been made to reduce this noise-like aspect by the process of de-speckling. These de-speckling operations lay a foundation for further application of information extraction in SAR images.

Let us recall the multiplicative speckle model, a common framework used to explain the stochastic behavior of data obtained with coherent illumination. The assumption made is that the observed pixel y in a SAR image is an outcome of the product of two independent random variables: one modeling the terrain backscatter x, which is generally considered real and positive; other modeling the speckle noise z, which could be complex (for complex-valued SAR images) or positive real (for amplitude and intensity SAR images). Again with the assumption of a fully developed speckle, the multiplicative noise model for SAR images is written as follows (Touzi [2002]):

$$y = x \cdot z. \tag{3.2}$$

With, y as the observed image pixel, x as the noise free pixel, i.e. the backscatter information, and z as the speckle noise. The objective of de-speckling is to obtain the most probable estimate $\hat{x}$ of the noise-free pixel x. Below we will discuss the de-speckling methods proposed along the years, the evolution of spatial content understanding and information extraction methodologies.

3.3.1.1 An overview on speckle reduction filters

One of the earliest methods for de-speckling was a multi-looking technique applied during the image formation. In this process several statistically independent looks of the same scene are averaged to reduce the speckle (Porcello *et al.* [1976]). Speckle is reduced but at the cost of reduction in ground resolution that is proportional to the number of looks L, so later many post-image formation de-speckling techniques have emerged.

The post-image formation de-speckling techniques can be classified into two groups, the non-Bayesian and the Bayesian filters. The non-Bayesian speckle filters consider only the local statistics of the scene computed with a continuous displacement of analyzing window. These kind of filters use the minimum mean squared error (MMSE) as an estimator. The Bayesian speckle reduction filters are based on models and an estimation of model's parameters, which determines how well the estimated data corresponds to the image data. The local adaptation is enhanced by the use of a *prior* model, which adapts to the image structure and texture in order to achieve de-speckling without loss in resolution. The local description of the structures estimated during this de-speckling operation forms a building block in the information extraction methods, analysis of which has been primarily carried out in this thesis.

Non-Bayesian speckle filters

The first category of non-Bayesian filters are very basic *simple filters* and do not take into account any speckle statistics. These are the **mean filter** and the **median filter**. No consideration is essentially given to exploit the most basic multiplicative model for speckle. Mean filtering or box averaging filtering is nothing but a multi-looking approach in spatial domain. Mean filtered gray-value for a given pixel is employed by taking the average value of a certain neighborhood of the given pixel, typically 5×5 or 7×7 pixels around a given pixel. This process will however suppress the speckle but also blur fine details such as edges and destroy the texture information at the same time. Median filter is applied over a neighborhood to a certain given pixel. The median value of this neighborhood will be a new gray-value for the given pixel. Median filter is a non-linear filter but does not assume any statistical model for the speckle as in a mean filter. But it preserves the edges and finer details better in the analyzing window, without introducing any new gray-values in the image (Rees & Satchell [1997]).

Under the second category of non-Bayesian filters comes the *statistical filters*, which take into account only the basic statistical properties of noise, i.e. mean and variance. Examples of statistical speckle filters are Kuan, Lee, Frost, Modified Lee, modified Frost, and EPOS filters etc.

The well known **Kuan filter** (Kuan *et al.* [1985]) and **Lee filter** (Lee [1980]) are adaptive filters, in which first a linear approximation is made for the multiplicative noise model, by first transforming the multiplicative noise model to a signal-dependent additive noise model, $y = x + (n-1)x$. Later a MMSE estimator in its simplest form is applied to obtain de-speckled pixel, i.e. $\mathbb{E}[(\hat{x} - x)^2]$ is to be minimized, where $\hat{x}$ is the estimate of actual noise-free pixel x. The speckle reduction filter can be written as:

$$\hat{x} = \mathcal{W}y + (1 - \mathcal{W})\mu_y. \tag{3.3}$$

The Kuan and Lee filters differ from each other with the formulation for the weighting function $\mathcal{W}$ computed from the local coefficients of variations, i.e. local mean μ_y and local variance σ_y^2 With L denoting the equivalent number of looks of the speckle noise, the weighting function $\mathcal{W}$

is given for Kuan and Lee filters as follows:

$$\mathcal{W}_{\text{Kuan}} = \frac{\sigma_x^2}{\sigma_x^2 + (\mu_y^2 + \sigma_x^2)/L}, \quad \mathcal{W}_{\text{Lee}} = \frac{\sigma_x^2}{\sigma_x^2 + \mu_y^2/L}. \tag{3.4}$$

Here, μ_y is the mean of the observed pixels in an analyzing window. The variance of the radar reflectivity σ_x^2 is a function of coefficients of variations μ_x and σ_x^2 for the observed pixel and equivalent number of looks L. The variance of radar reflectivity is derived as $\sigma_x^2 = (L\sigma_y^2 - \mu_y^2)/(L+1)$. It can be seen from the equation 3.4 that the Lee filter is a special case of the Kuan filter because a linear approximation is made for the multiplicative noise model.

The **Frost filter** (Frost *et al.* [1982]) differs from the Lee and Kuan filters with respect that the scene reflectivity is estimated by convolving the observed image with the impulse response of the SAR system (Shi & Fung [1994]). With the assumption that scene reflectivity model obeys an auto-regressive process, the simplified formulation for the impulse response can be written as:

$$m(r) = \mathcal{K}_1 \exp\left(-\mathcal{K}\frac{\sigma_y^2}{\mu_y^2} \mid r \mid\right). \tag{3.5}$$

Here, $\mathcal{K}$ is the filter parameter, that controls the de-ramping rate of the impulse parameters, thus determines the strength of the filtering. $\mathcal{K}_1$ is a normalization factor and $\mid r \mid$ is the radial distance from the center pixel to be de-speckled. It can be considered as an adaptive Wiener filter, which is the optimal filter in the sense of MMSE for stationary signals corrupted by additive noise. But after many approximation in the impulse response as presented in the equation 3.5, the overall performance is comparable to most other statistical filters.

In Lopes *et al.* [1990b], an enhanced formulation for Lee and Frost filters were presented. The underlying idea in **modified Lee filter** and **modified Frost filter** was to divide an image into three classes: homogeneous regions, heterogeneous regions and isolated point targets. Speckle reduction in homogeneous regions was proposed to be carried out by a simple low pass filter, while isolated point target regions were suggested to be preserved as it is. Now the speckle reduction filter is forced to be applied only to the heterogeneous regions. This is achieved by exaggerating the filter response by introducing a hyperbolic function to satisfy the requirement that the more heterogeneous the region is, the less it has to be smoothed (Lopes *et al.* [1990b]). It is similar to the observations made in Lee [1980] that images containing multiplicative noise have the characteristic that the brighter the area the noisier it is.

Another method to guarantee an improved edge-preserving capabilities is called **EPOS (Edge Preserving Optimized Speckle-filter)** (Hagg & Sties [1994]. EPOS filter is also based upon the statistical properties of speckle noise. As an implementation point of view, the image is divided into 8 even sized mutually exclusive triangular sub-windows. The assumption in preserving the edges is that the relative standard deviation $\sigma_y^2(i) = (1, 2, \cdots, 8)$ in these sub-windows remain nearly constant in the homogeneous areas, while it increases in the presence of edges. Thus knowledge about the standard deviation σ_y^2 will inform us about the homogeneous regions, heterogeneous regions or point targets in a given area. The mean of the most homogeneous region is considered as the estimate of the noise free pixel $\hat{x}$. One disadvantage is that even though this filter preserves the edges, the texture is absolutely not preserved.

It is evident from the presented discussion that the mentioned adaptive non-Bayesian filter will provide best accuracy in case of the Gaussian distributed noise. But this is certainly not the case for SAR images (Oliver & Quegan [1998]). The other disadvantage is that no matter how the multiplicative model is followed, only the statistical considerations on speckle are accounted for without considering a model for the reflectivity of the scene. These disadvantages lead to the concept of Bayesian filters.

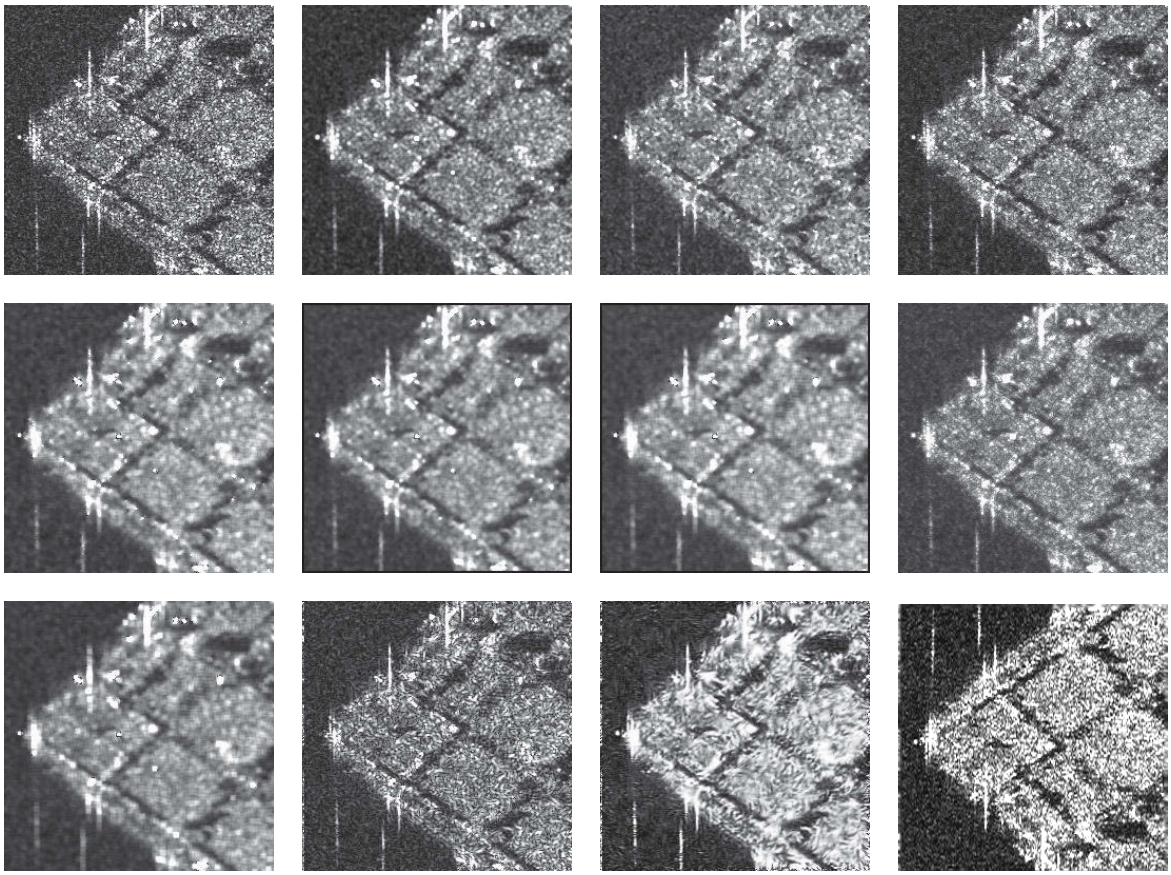


Figure 3.10: Illustration of visual performance of various non-Bayesian and Bayesian despeckling algorithm on very-high-resolution spotlight mode TerraSAR-X image. First row from left to right: original SAR images, mean filtered image, median filtered image, Lee filtered image. Second row from left to right: Enhanced Lee filtered images, Frost filtered image, Enhanced Frost filtered image, Kuan filtered image. Third row from left to right (maximum a posteriori (MAP) filters): Gamma MAP filtered image, MBD using GMRF *prior*, MBD using ABM *prior*, MBD using GMRF *prior* in wavelet domain. Analyzing window size used for speckle suppression: 5×5 pixels.

Bayesian speckle filters

The important aspect of Bayesian speckle filters is the statistical understanding of the reflectivity of the scene, along with the statistical understanding of the noise model, with the application of maximum *a posteriori* (MAP) [1]. The MAP estimate is obtained by maximizing the Bayes criterion (the MAP pdf) with respect to the *posterior* pdf $p(x \mid y, \boldsymbol{\theta}) = p(y \mid x)p(x \mid \boldsymbol{\theta})/p(y \mid \boldsymbol{\theta})$, where $p(y \mid x)$ corresponds to the pdf of the speckle (called *likelihood*) which is known to be a Gamma distribution in the case of fully developed speckle and when speckle samples are assumed to be independent of each other. A statistical model $p(x \mid \boldsymbol{\theta})$ with model parameters $\boldsymbol{\theta}$, concerning the underlying scene reflectivity is called *prior* model. The task in MAP is to search for $\hat{x}$ which would maximize the *posterior* pdf, i.e. $\hat{x} = \arg\max_x p(y)$.

The application of maximum *a posteriori* to de-speckling was first proposed in Kuan *et al.* [1987] with the model for scene reflectivity, i.e. *prior* pdf, to be Gaussian distributed. Later a modified MAP filter was proposed in Lopes *et al.* [1990a] by selecting the model for *prior*

[1] The Bayesian inference will be covered in chapter 4 in detail.

as Gamma distribution as well. This filter is denoted as Gamma-Gamma MAP filter and is a widely used de-speckling filter with following MAP estimate equation (Lopes *et al.* [1990a]):

$$\hat{x}_{\mathrm{MAP}} = \frac{(\alpha - L - 1)\mu_y + \sqrt{(\alpha - L - 1)^2\mu_y^2 + 4\alpha L y \mu_y}}{2\alpha}, \quad \text{with } \alpha = (L+1)/(L\sigma_y^2/\mu_y^2 - 1). \tag{3.6}$$

Here, L is the equivalent number of looks, and μ_y and σ^2 are the mean and variance of the observed pixels in the analyzing window. This filter is more elaborate than the non-Bayesian filters but as the parameter estimation is depending upon local mean and local variance only, it is still not able to cater for the complexity of the scene reflectivity.

The demand for encapsulating more information than mean and variance was felt for more complex parametric models (such as Markov Random Field and Gibbs Random Fields based models) as scene reflectivity to describe the spatial structures. The use of Gibbs Random Fields (GRFs) models as a prior model of the scene for de-speckling SAR images was introduced in Geman & Geman [1984], by encapsulating more information than just mean and variance in order to describe the spatial structures in the images. Use of Gauss-Markov random fields (GMRF) and autobinomial (ABM) model as *prior*, both belonging to the category of Gibbs Random Fields was introduced in Walessa & Datcu [2000] and Hebar *et al.* [2009], respectively, for model based de-speckling (MBD). In both cases a neighborhood system around a central pixel was used in order to give the value of the central analyzed pixel considering its spatial relations with its neighbors.

Figure 3.10 illustrates only the visual de-speckling performance of a few of the discussed non-Bayesian and Bayesian approaches. A quantitative analysis of de-speckling approaches is not the goal of this thesis.

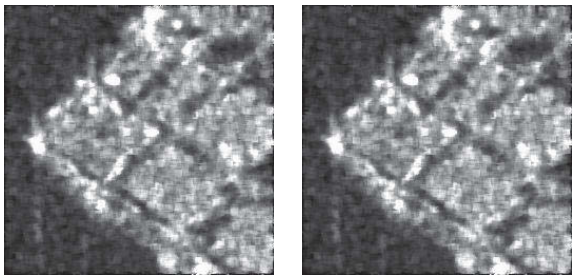
From de-speckling to information extraction

In the beginning of this section we mentioned that a very compact intrinsic feature descriptor is needed to describe the content of image. The underlying principle of the presented de-speckling method in Bayesian framework is the assumption of *prior* pdf as scene reflectivity to describe the spatial structures and spatial content understanding in a broader sense. The model parameters of assumed *prior* pdf can be used as an element of the parametric feature descriptors describing the image content. Motivated from the impressive conclusion about the strength of complex-valued SAR data in MSLD method, we also tried to extend the GMRF in the complex-valued space (Singh *et al.* [2009], Soccorsi [2010]). Chapter 4 is dedicated to the guidelines in the formation of model based feature descriptors.

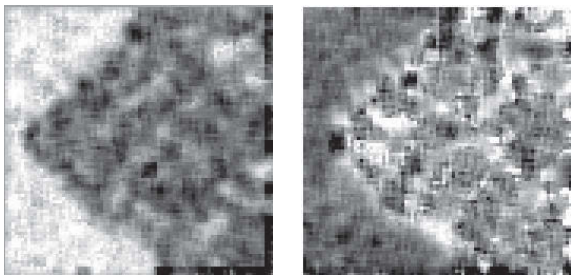
3.3.1.2 Feature descriptor performance

To summarize the overview on spatial content understanding, we considered mainly two categories of feature descriptors. One based on the time-frequency analysis and the other based on the model parameters. Without going into too many technical details, which will be explained in the coming chapters, in the figure 3.11, some examples of the features descriptors generated using some selected methods from both categories are presented.

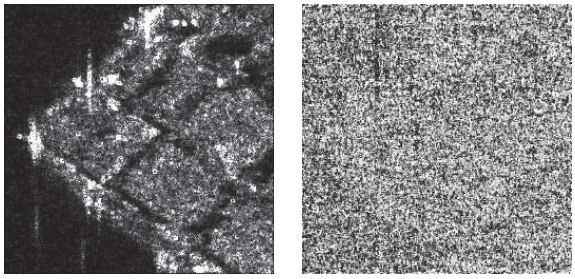
At this stage its pertinent to mention that one of the main objective of this thesis is to qualitatively examine the suitability of selected feature descriptors for two applications areas of SAR images, namely, pixel-based classification and patch-oriented image categorization. In this thesis, we will restrict ourselves to the eight methods, which are mentioned in the figure 3.11.



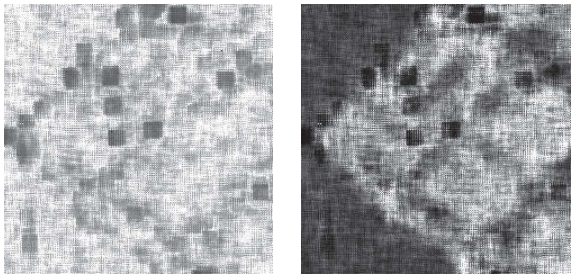
(a) GMRF as *prior* pdf.



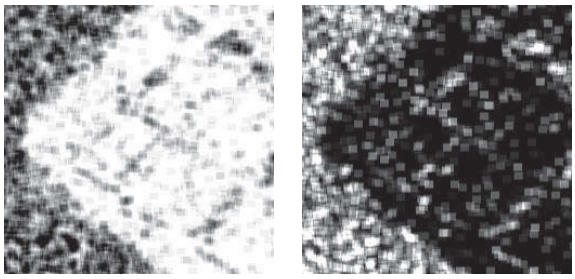
(b) ABM as *prior* pdf.



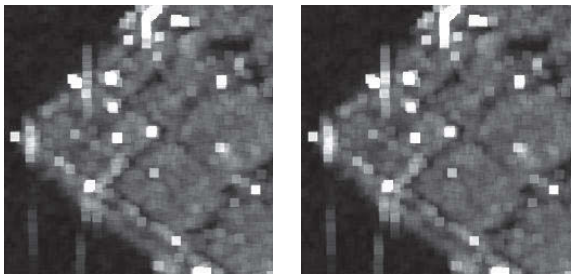
(c) GMRF as *prior* for complex-valued SAR images.



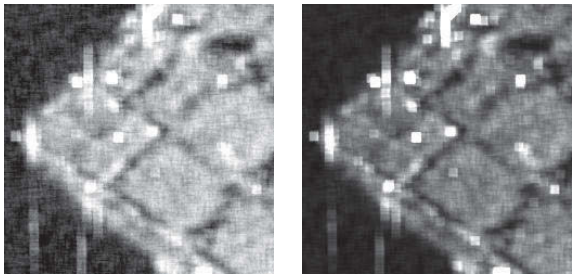
(d) GGD parameters for Wavelet coefficients.



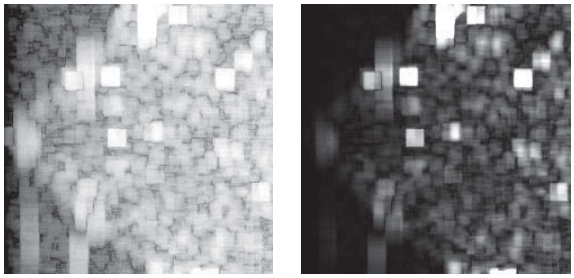
(e) Gray-level co-occurrence matrices.



(f) Non-linear spectral features.



(g) The fractional Fourier transform based feature.

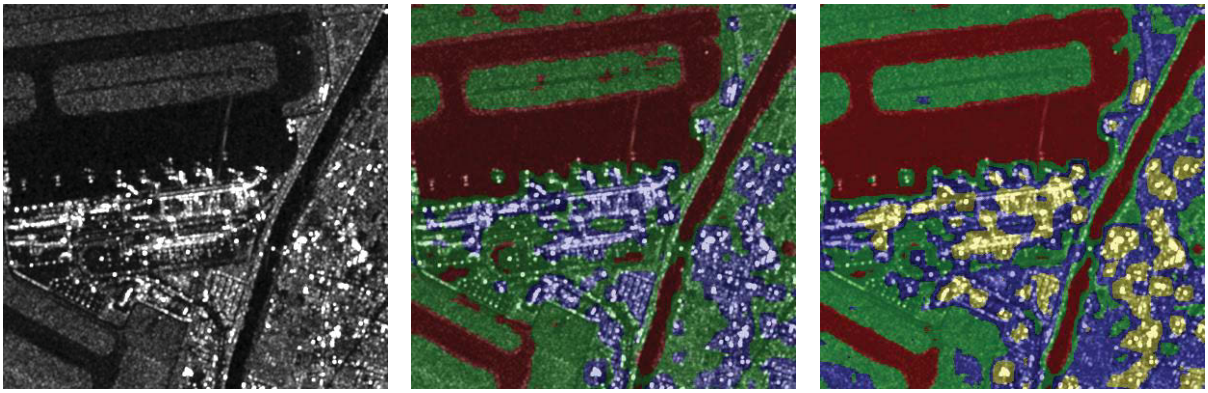


(h) Gabor filter banks based features.

Figure 3.11: Information extraction using several methods. (a) Gauss-Markov random fields as *prior* pdf for MAP estimation at model order 3, left- model variance and right- first model parameter, θ^1. (b) Autobinomial model as *prior* pdf for MAP estimation at model order 3, left- model variance and right- first model parameter, θ^1. (c) Gauss-Markov random fields as *prior* pdf in complex-valued SAR images for MAP estimation at model order 3, left- model variance and right- first model parameter, θ^1. (d) Generalized Gaussian distribution (GGD) model parameters for the wavelet coefficients, number of decomposition level- 3, left- shape parameter for H1 and right- scale parameter for H1. (e) The gray level co-occurrence based features at offset=1 and orientation = 90°, left- variance and right- contrast. (f) Non-linear features based on the spectral analysis, left- spectral centroid in azimuth and right- spectral centroid in range. (g) The fractional Fourier transform based features at transform angle $\alpha = 45°$, left- first moment and right- first logarithmic cumulant. (h) The Gabor filter banks based features at stage-1 and orientation-1, left- first moment and right- first logarithmic cumulant. Analyzing window size used for sliding window operation for all but (d) and (h) was 7×7 pixels, with step size of 1 pixel. For the multi-scale case of (d) and (h), analyzing window of 16×16 pixels size was used.

3.4 From pixel-based classification to patch-oriented image categorization

One among the earliest applications of SAR images, after the launch of Seasat and the space shuttle mission was land-use segmentation and classification. With such low-resolution images, the interest was always in global criteria. A problem statement in target detection was to detect global classes, e.g. different kinds of vegetation, urban areas or water bodies etc. So the end result usually was the image segmented into different classes / clusters, with supervised or unsupervised means. Over a period of time and with the advent of other low and medium resolution SAR satellites, many advanced algorithms have been developed for land use applications, due to the visibility of so called *texture*. We noticed in the examples presented in the figure 3.5, that the importance of texture was still relevant, and so is the pixel-based classification. An example has been presented in the figure 3.12. The image under consideration in figure 3.12a is



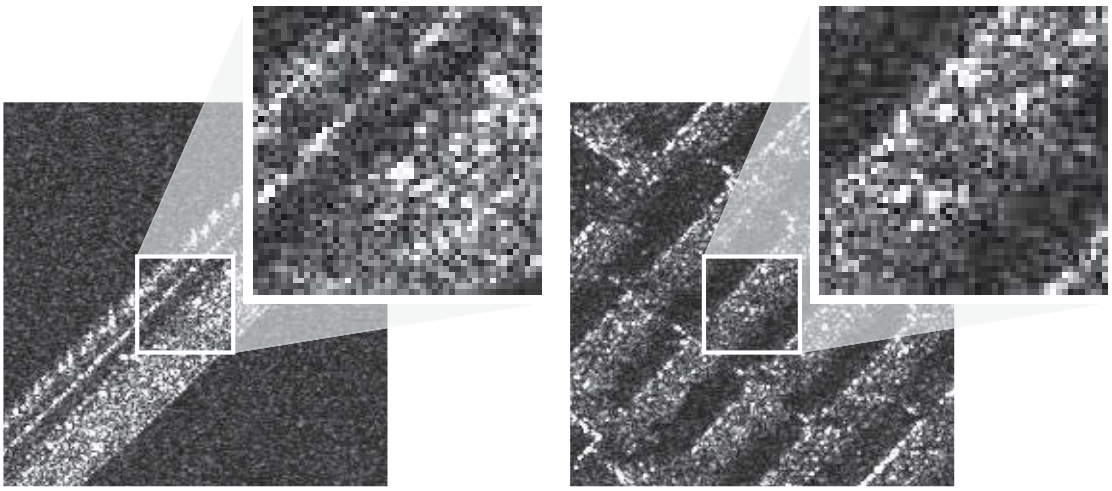
(a) SAR image. (b) Classification map, 3 classes. (c) Classification map, 4 classes.

Figure 3.12: Example of pixel-based classification to generate a classification map. (a) SAR image of size 400×400 pixels. Texture features have been synthesized using Gauss-Markov random fields as *prior* pdf at model order 3. Analyzing window used for texture parameters is 16×16. (b) Unsupervised k-means classification into 3 classes. (c) Unsupervised k-means classification into 4 classes.

an image of size 400×400 pixels from the stripmap mode TerraSAR-X with 3 meter resolution. We have computed a parametric feature descriptor with an analyzing window of 16×16 pixels, based on the *prior* Gauss-Markov random fields, with model order 3 and the speckle is modeled as Gamma distribution. Cluster map in 3 and 4 clusters using k-means unsupervised clustering method have been presented in the figures 3.12b and 3.12c, respectively. We notice that in the figure 3.12b, the 3 global clusters corresponds to - airstrip (*red*), open fields (*green*), and man-made objects (*blue*). In the figure 3.12c, the 4 global clusters corresponds to: airstrip (*red*), open fields (*green*), man-made objects (*blue*), strong scatterers in man-made objects regions (*yellow*). There are many regions that does not match our given semantic meanings, such as the water channel which is in the same class as airstrip and so on. This was the case of relatively lower resolution, which was a perfectly good example of application area of lower resolution images for large scale land-use mapping, monitoring of crops and urban area expansion, mapping of disaster activities etc.

In the figure 3.5, we see a different perspective of the metric-resolution SAR images, where instead of texture, objects are visible. This makes it possible to analyze individual targets. But with individual target analysis using a small analyzing window in high and very-high-resolution

SAR comes the problem of similarity in the backscatter from various scatterers, which otherwise belongs to different categories. This statement can be understood by referring to the figure 3.13 which depicts two different target under consideration. Figure 3.13a is a part of an over-water bridge and the figure 3.13b is a part of an urban area with relatively large building blocks.



(a) Bridge with a part of bridge in inset

(b) Building with a part of building in inset

Figure 3.13: Importance of contextual information in case of metric-resolution SAR images. Without the contextual information, a small part (in inset) of (a) bridge and (b) building structure will *look* similar. A larger analyzing window incorporate relevant contextual information, which makes both the objects comprehensible and provides a semantic meaning to the objects.

Considering a smaller analyzing window than the main object in both the cases (shown in inset in both cases), they *look* very similar and even exhibits similar backscattering behavior. But the meaning of the main object changes with the way even similar behaving scatterers are arranged and largely on the contextual information between them. So the presence of water on the upper and lower part of the analyzing window in the figure 3.13a will aid us to identify it as an over-water bridge, and the regular arrangement of same objects as seen in the analyzing window in the figure 3.13b will aid us to identify it as most likely an urban area. So the semantic meaning of the objects will be largely different with different arrangement of same objects. Thus in high-resolution and very-high-resolution SAR images, our focus is now on the patch-oriented image analysis, which might contain similar kind of scatterers on a global scale but differ significantly on a small scale due to the complicated arrangement of these scatterers in a patch.

In high-resolution and very-high-resolution SAR images, we are now more importantly focused on patch-oriented image categorization in contrast to the pixel-based classification in low-resolution SAR images. By patch-oriented image categorization, we mean that our objective is not to classify some regions of the image, but the whole image-patch into an object category. The patch-oriented image categorization can be considered as a extension of the classical automatic target recognition (ATR) technique. The goal of ATR is to detect and recognize objects, such as tanks, in images produced by a laser, a radar, a SAR, an infra-red or a video camera (Novak *et al.* [1997]). Generally in ATR we deal with the targets of interest located on a homogeneous background like vegetation or bare land. Prior models of the isolated targets are generally available in the data set to ease the recognition of the object. Examples for ATR are

shown in the figures 3.14(a,b,c). Here the objective could be to detect and recognize the type of tank, electric pylon and aircrafts, respectively.

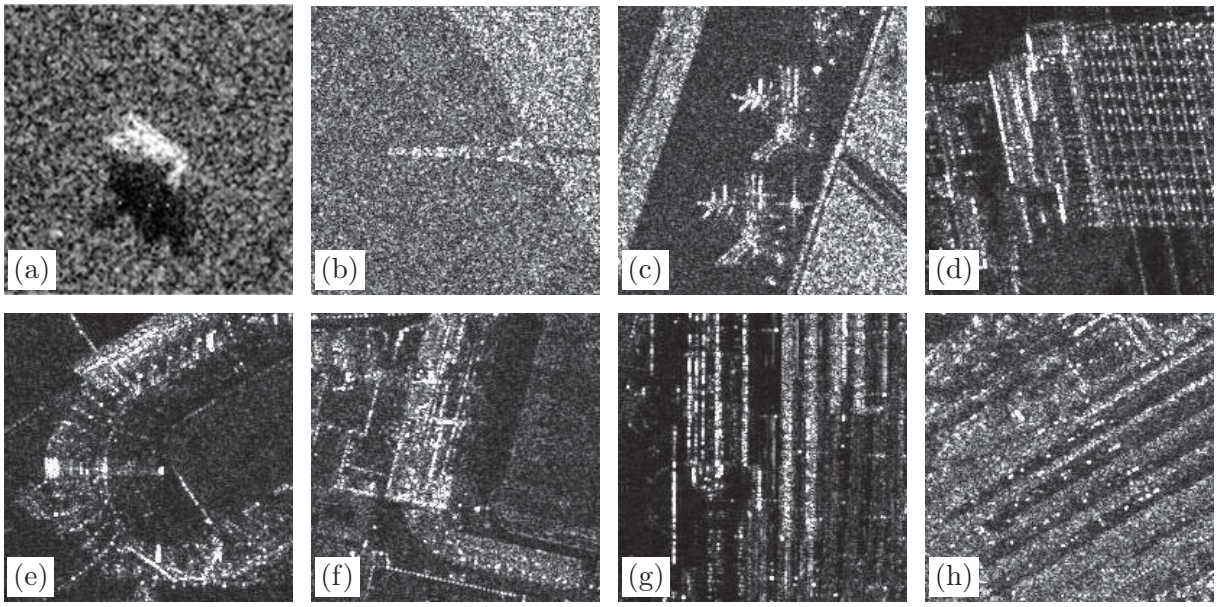


Figure 3.14: Examples from high-resolution SAR images, (a) size is 128×128 pixels and (b-h) size is 200×200 pixels. (a) a typical example of ATR image from MSTAR database (Ross *et al.* [1998]): T-72 battle tank, (b,c) potential candidates for ATR from a very-high-resolution spotlight mode of TerraSAR-X, electric pylon and aircrafts respectively. (c-h) candidates for patch-oriented image categorization, aircraft-stands, skyscrapers, stadium, stadium, train-lines, and train-lines respectively.

As an extension to ATR, we are now more interested in the categories which are formed with the complex arrangement of objects, not necessarily present on a homogeneous background. Categories will be defined as groups of objects sharing the same semantic. For example, an objective in patch-oriented image categorization will be to recognize the figure 3.14(c) in the category of aircraft-stands, the figure 3.14(d) in the category of high-rise buildings, the figures 3.14(e,f) in the category of sports-stadiums and the figures 3.14(g,h) in the category of train-stations. Such an approach will find an application in the content-based image retrieval (CBIR) systems for SAR images. CBIR system are already widely applicable in optical remote sensing images (Shyu *et al.* [2007]) as well as in multimedia (Smeulders *et al.* [2000]). The task will be to retrieve all the relevant image patches from a huge repository of image patches based upon the positive and negative examples given to train a classifier. Positive example will belong to the desired category and negative examples to all other categories. With the use of a look-up table with meta-data of the retrieved patches, further post-retrieval operations can be performed.

However, while the main focus is on giving a semantic meaning to images based on the basis of the arrangement of scatterers, the issue of texture in homogeneous area should not be ignored. Thus a general framework of the application area can be defined on the basis of image content, as shown in the figure 3.15. The number of total global classes has been found limited to $5-10$ in every scenario, whereas its very easy to find more than 20 patch-oriented categories even in a single sufficiently-diverse full scene of spotlight mode TerraSAR-X image.

As each application, pixel-based classification and patch-oriented image categorization, is more important than the other depending upon the user requirement, we will consider both cases for the performance analysis of the considered feature descriptors.

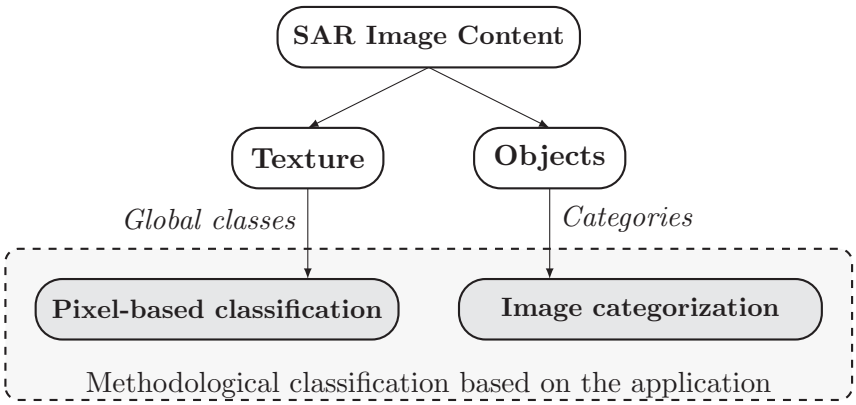


Figure 3.15: Application framework based on the image content. It is suitable to carry out land-cover pixel-based classification with prominent texture, whereas with objects visibility, patch-oriented SAR image categorization is more suitable.

3.4.1 Proposed methodological classification for performance analysis

In this last section of the chapter, we propose a methodological classification, initially presented in Singh & Datcu [2012a], which will form a basis of some of the conclusion drawn in this thesis. In section 3.3, we noticed that some of the feature descriptor were the parameters based on model assumed to fit the SAR images. Basically, all the feature descriptors were based on Bayesian inference. However, a good fit of a particular model requires meeting many preliminary assumptions to satisfy the multiplicative model of SAR speckle. This category can be called the **parametric approaches** (or model-driven approaches). The other class of feature descriptor was based on simple statistical measure, such as GLCM based features, or simple statistical measures computed on the transformed SAR images, such as non-linear spectral features, the FrFT based features, and Gabor filter banks based features. No assumption is required to be met in these scenarios and we can call these features to be based on the **nonparametric approaches**.

This discussion leads to a methodological classification under parametric and nonparametric approaches in images and images within a transformation space, as presented in the upper part of the figure 3.16. No matter what kind of methodology we are adapting for generating a feature descriptor, they can be put under this methodological classification. The success and performance of a particular feature descriptor will depend upon the type of application and image content of the image under consideration.

In this thesis, apart from suggesting a new nonparametric methodology based on the FrFT, we will provide a quantitative analysis of selected methods presented in the lower part in the figure 3.16.

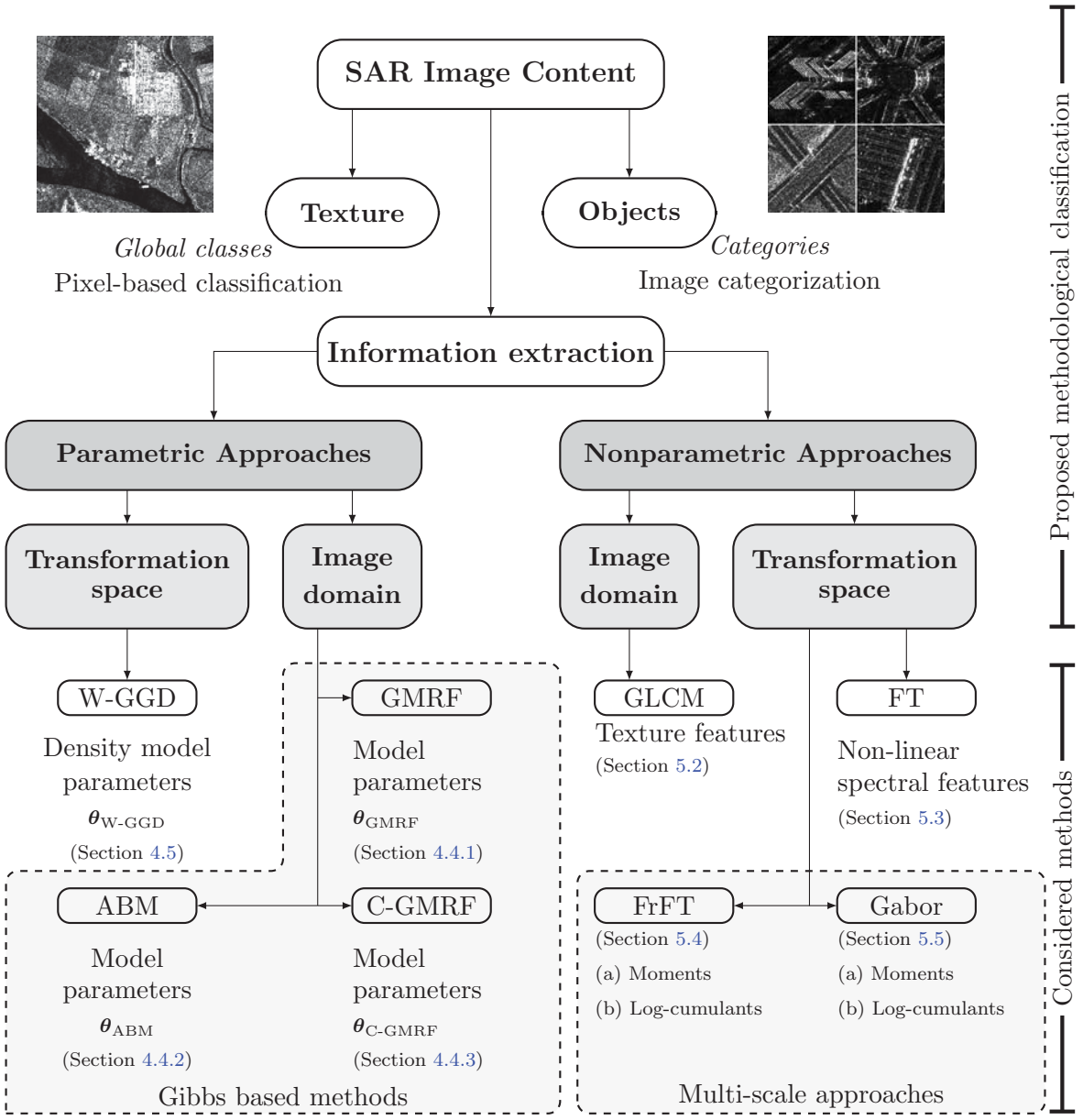


Figure 3.16: Proposed methodological classification under *parametric* and *nonparametric* approaches in *image* and *image within a transformation space* Singh & Datcu [2012a]. This methodological classification comes into existence because of the possibility of discovering a high diversity of texture as well as objects in the high-resolution SAR images. In this thesis, we will mainly provide a detailed qualitative and quantitative analysis of various state-of-the-art and few proposed methods for pixel-based and patch-oriented image categorization applications. Acronyms- GMRF: Gauss-Markov random fields as *prior* pdf, ABM: autobinomial model as *prior* pdf, C-GMRF: Gauss-Markov random fields as *prior* pdf for complex-valued SAR images, W-GGD: generalized Gaussian distribution parameters for the wavelet coefficients, GLCM: gray level co-occurrence matrix, FT: Fourier transform, FrFT: The fractional Fourier transform, Gabor: Gabor filter banks.

3.5 Summary

The following points have been discussed in this chapter.

- The ever increasing resolution from recent SAR satellite missions are changing the problem statement in SAR image understanding. While the concern in earlier low-resolution image was to identify global classes such as natural areas, urban area and water bodies etc., in high and very-high-resolution images we are more concerned with patch-oriented image categorization.

- The spatial content understanding in SAR images depends upon the available resolution, which also governs the underlying applications area. Even though many high and very-high-resolution SAR satellites are planned for the future, satellite missions such as Sentinel-1 will keep us involved in developing further algorithms in global classification, by analysis and characterizing the texture especially for large scale mapping for land application, vegetation and sea-ice monitoring.

- In high and very-high-resolution images, the information available is sometimes sufficient enough to even carry out individual target analysis. Based on the concept of azimuth splitting, we proposed an approach of visual exploration of targets in complex-valued SAR images, termed as multiple sublook decomposition (MSLD). This tool provides us an insight into the complex-valued spectrum of SAR images to reveal some information which is otherwise hidden in the spatial domain.

- Motivated by the information retrieved from complex-valued spectrum in the MSLD approach, we proposed the use of a generalization of the Fourier transform (the fractional Fourier transform) for a broader spatial content understanding, use of which was limited to moving target analysis until now.

- In order to develop and understand methodologies for spatial content understanding based on the statistical analysis, it is pertinent to study the speckle and speckle reduction methods. Historical overview of de-speckling has been covered which eventually leads to the application of the Bayesian approach for SAR image understanding.

- Based on the visibility of texture or even more detailed information, we proposed a methodological classification as a guideline for adapting certain type of methodology for certain kind of application. This leads us to a performance analysis strategy for various spatial content understanding methods, which will be discussed in detail in the remainder of the thesis.

Based on the methodological classification proposed in this chapter, the next part of this thesis, which is Part-II, will deal with the implementation issues of the selected and proposed parametric and nonparametric approaches for spatial content understanding. Theoretical understanding and selective results will be presented in Part-II. The parametric approaches are discussed in chapter 4, whereas nonparametric approaches are discussed in the chapter 5, of Part-II.

Part II

Information Extraction in SAR Images

Chapter 4

Information extraction in SAR images: parametric approaches

The objective of this chapter is to discuss the parametric approaches for information extraction in SAR images, which can be further classified into texture parameter estimation in the image domain or density parameter estimation for the image within a transformation space such as wavelet domain. For both the cases, a particular model is postulated for the data and later the parameters of the assumed model are considered as the descriptor for the image.

In both the mentioned cases, the probability theory plays an important role, thus we discuss the basics of probability theory in the first section. For the texture parameter estimation, usually Bayesian inference is applied to SAR images as a consequence of the multiplicative noise model for speckle. In second section, we present the theoretical understanding of Bayesian inference, along with the first level of Bayesian inference for model fitting and parameter estimation, and second level of Bayesian inference for model selection.

In the third section, we discuss the applicability of Bayesian inference to the SAR images, by introducing the concept of neighborhood systems, cliques and potential functions for prior pdf. The usual choice for the prior pdf are the Gibbs based models. Two Gibbs based prior pdfs, namely, Gauss-Markov random fields and autobinomial model will be discussed in detail. Guidelines for parameter estimation using maximum a posteriori (MAP) along with preliminary results will be presented. We also propose the applicability of the Gauss-Markov random fields to the complex-valued SAR data and parameter estimation methodology for this case is suggested.

In the last section, we briefly discuss a parametric approach applied in the wavelet domain by characterizing the wavelet coefficients with generalized Gaussian distribution (GGD). The parameters of the GGD for wavelet sub-bands can be used as the image descriptor. We will close this chapter with a short summary in the last section.

4.1 Probability and Bayesian inference

As the name suggest, the parametric approaches for information extraction in SAR images are based on the density or model-parameter estimation. Before going into the discussion on the models and model-parameter estimation, it is important to briefly discuss the probability theory. Probability theory is essential for the formulation of parametric models. During this discussion we will also introduce the mathematical notations used in this thesis to describe the probability measures.

Probability theory is nothing but common sense reduced to calculation (Pierre-Simon Laplace, 1812). The interpretation of the probability theory is generally divided into two notions. The first one is the **frequentist** probability notion and the second one is the Bayesian notion of probability. In the frequentist notion of probability, the probability represents long run frequencies of events. Using this frequency interpretation, the probability $P(A)$ of an event A is given as follows:

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n} \tag{4.1}$$

with, n_A as the number of occurrences of the event A and n as the number of trials. For example, flipping a coin many times i.e. many trials, we expect the occurrence of facing heads or tails about half the time. The other interpretation of the probability notion is called the **Bayesian** interpretation of probability, where probability is used to quantify our uncertainty about something; hence it is fundamentally related to information rather than repeated trials (Jaynes [2003]). In the Bayesian approach, the likelihood of occurrence of facing heads or tails on the next trial of tossing a coin is equal.

- **Fundamental axioms in the theory of probability**

The probability $P(A)$ of an event A, defined on the probability sample space Ω, is usually chosen in such a way that it satisfies the following three Kolmogorov axioms:

1. Positivity: $0 \leq P(A) \leq 1$ for every event A.
2. Certainty: $P(\Omega) = 1$.
3. Additive rule: $P(A \cup B) = P(A) + P(B)$, provided events A and B are mutually exclusive, i.e. $P(A \cap B) = 0$.

Here $P(A) = 0$ means that the event will definitely never occur, i.e. the probability of an impossible event is 0. $P(A) = 1$ means that the event will certainly occur. Similarly we can also define $P(\bar{A})$ to denote the probability of an event not A. Thus from the positivity and certainty axiom, it can be be established that $P(A) = 1 - P(\bar{A})$. All the subsequent rules and theorem are the direct or indirect derivations of the above mentioned three fundamental axioms.

- **Discrete random variables**

The above discussion mainly dealt with the binary events, and can be extended by defining a discrete random variable X, which can take any value in the probability sample space Ω. I.e. $X : \Omega \to \mathbb{R}$ is a discrete random variable defined in space Ω. We can also denote the probability of this event $X = x$ by $P(X = x)$. In the subsequent sections we will also sometimes use a short notation $p(x)$ for $P(X = x)$. In case of discrete random variables, $p(x)$ is also denoted as the probability mass function (pmf). The pmf also satisfies the fundamental properties: $0 \leq p(x) \leq 1$ and $\sum_{x\in\Omega} p(x) = 1$.

- **Conditional probability and Bayes' theorem**

Apart from the three axioms, we can also define the product rule, which describes the joint probability of two events A and B as $P(A, B) = P(A \mid B)P(B)$. The product rule is also known as the *total probability theorem*. Here, the probability $P(A \mid B)$ is the conditional probability of an event A, given the event B is true, which can be subsequently written as follows:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}, \quad \text{provided } P(B) > 0. \tag{4.2}$$

The frequency interpretation of the conditional probability can also be derived. Denoting n_A, n_B and $n_{A,B}$ as the number of occurrences of the events A, B and (A, B), we have $P(A) = n_A/n$, $P(B) = n_B/n$, and $P(A, B) = n_{A,B}/n$. Thus the conditional probability $P(A \mid B)$ of an event A, given the event B is true, is given as follows:

$$P(A \mid B) = \frac{P(A, B)}{P(B)} = \frac{n_{A,B}/n}{n_B/n} = \frac{n_{A,B}}{n_B}. \tag{4.3}$$

This above formulation means that if we discard all trials in which the event B did not occur and we retain only the subsequence of trials in which B occurred, then $P(A \mid B)$ equals the relative frequency of occurrence of $n_{A,B}/n_B$ of the event A in that subsequence (Papoulis [1991]).

The three axioms presented earlier form the basic algebra of the probability theory. As mentioned earlier also that the other theorems can be derived from these fundamental axioms. One important theorem obtained by combining the product and additive rule with conditional probability can be defined as follows (Bayes [1763]):

$$P(X = x \mid Y = y) = \frac{P(Y = y \mid X = x)P(X = x)}{P(Y = y)}, \tag{4.4}$$

which describes the probability of an event X, given some data Y. The formal names for each term in the Bayes' Theorem are well defined, which will also be followed in the remainder of the thesis. The left-hand side term $P(X = x \mid Y = y)$, called the *posterior probability*, is proportional to the probability of the observation given the hypothesis $P(Y = y \mid X = x)$, called *likelihood function*, multiplied with the *prior information* $P(X = x)$. The denominator term on the right-hand side, i.e. $P(Y = y)$, is called the *evidence*, which is in the first level of Bayesian inference treated as only a normalization constant. The Bayes' theorem can be written in words as follows:

$$posterior = \frac{likelihood \times prior}{evidence} \tag{4.5}$$

Bayesian inference applied to the information extraction in SAR images, based on Bayes' theorem allows us to select the most probable model, i.e. *prior information* for the SAR images.

- **Continuous random variables and density functions**

So far, we were only concerned in our reasoning with the uncertain discrete random variables. A similar reasoning can be extended to the uncertain continuous random variables. Bayes' theorem is very well applicable to continuous random variables as well. In this section, we are interested in introducing the terms cumulative density function (cdf) and probability density function (pdf) for a continuous random variables X. For this purpose, let us compute the probability for a continuous random variable X, such that it lies between any interval $a \leq X \leq b$.

Defining the events, $A = (X \leq a)$, $B = (X \leq b)$, and $W = (X \in (a, b\,])$, using the fundamental axioms and assuming A and W as mutually exclusive, we obtain $p(B) = p(A)p(W)$, and hence:

$$p(W) = p(B) - p(A) \tag{4.6}$$

Denoting $F(x_1) \stackrel{\Delta}{=} p(X \leq x_1)$ as the cdf, which is a monotonically increasing function, it can be written as follows:

$$p(X \in (a, b\,]) = F(b) - F(a) \tag{4.7}$$

The derivative of the cdf $F(x)$ is called the pdf or simply the density function $f(x) = \frac{d}{dx}F(x)$ of the random variable x, assuming that this derivative exists. Given a pdf, the probability of a continuous random variable in a finite interval can be computed as follows:

$$P(X \in (a, b\,]) = \int_a^b f(x)dx. \tag{4.8}$$

For an infinitesimal small interval, we obtain:

$$P(X \in (x, x + \Delta x\,]) \approx p(X = x)dy. \tag{4.9}$$

The most important property of a distribution are it's moments. The first moment is called the mean, or expected value $\mathbb{E}[X]$, denoted by μ. The second moment, variance $\mathbb{E}[(X - \mu)^2]$, denoted with σ^2 is a measure of the spread of the distribution. If we assume a specific functional form for the densities, i.e. a specific distribution $p(X = x \mid \boldsymbol{\theta})$ for the data X, with $\boldsymbol{\theta}$ as the model parameter vector; the task of parameter estimation is to estimate the optimum model parameters $\boldsymbol{\theta} = \theta_1, \theta_2, \cdots, \theta_p$, in order to obtain a good fit of the assumed model to the data.

In literature, notation p is used interchangeably for probability as well as for probability density function, for both the discrete and continuous case. For the sake of simplicity, we will also follow the same convention and meaning of p can be understood as per the context.

4.2 Bayesian inference preliminaries

4.2.1 Bayesian inference, level I: model fitting and parameter estimation

Denoting $p(\cdot)$ as probability, as well as the probability density function and assuming a certain *prior model* $(X = x)$ for a data, we can rewrite the Bayes' theorem as follows:

$$p(X = x \mid Y = y) = \frac{p(Y = y \mid X = x)p(X = x)}{p(Y = y)}, \tag{4.10}$$

The task of parameter estimation is to infer the parameter value $\hat{x}$ from the observation y, i.e. the task is to infer a variable X conditioned by another variable Y. There can be many ways to estimate the value of $\hat{x}$. However, in this discussion we restrict ourselves only with the maximum likelihood estimator (MLE) and maximum *a posteriori* (MAP) estimators for finding the value of $\hat{x}$. To determine the most probable value for X, the task is to search for $\hat{x}$ which would maximize the *posterior* probability $P(X = x \mid Y = y)$, i.e.:

$$\hat{x} = \arg\max_x p(X = x \mid Y = y). \tag{4.11}$$

In case we do not have any further information about the *prior model* M, $p(M)$ is taken to be a uniform distribution. So we obtain an estimate for $\hat{x}$, as follows:

$$\hat{x}_{\text{MLE}} = \arg\max_x p(Y = y \mid X = x) = y, \tag{4.12}$$

which is called the MLE of X. For MLE, only the knowledge or assumptions about the *likelihood function* is required (Kreyszig [1979]). However, if we have further knowledge about the *prior model*, the task of parameter estimation is to maximize the *posterior probability*, similar to the equation 4.12:

$$\hat{x}_{\text{MAP}} = \arg\max_x p(X = x \mid Y = y) = \arg\max_x p(Y = y \mid X = x)p(M), \tag{4.13}$$

which is called the maximum *a posteriori* (MAP) estimate, also termed as the first level of Bayesian inference. We can state that the MLE is a special case of MAP estimator when the *prior model* is assumed to be uniform distribution.

The first level of Bayesian inference, the problem of MAP estimation has been criticized because of the choice of the correct *prior distribution*, i.e. correct model M, which is compensated with the second level of inference (the model selection). On a philosophical view of the discussion, it has also been said that all models are wrong, but some models are useful (Box & Draper [1987]). In this thesis, we will also try to establish which models are useful in which circumstances for information extraction the SAR images.

4.2.2 Bayesian inference, level II: model comparison and selection

We noticed in section 4.2.1, the first level of Bayesian inference employs the use of a *prior distribution*, choice of which is a very critical issue. The second level of Bayesian inference, i.e. the model selection process provides a framework to search for the most plausible model from the different alternative models. Here, the *evidence* term plays a fundamental role, which has been assumed just as a normalization factor in the parameter estimation or the first level of Bayesian inference. The *evidence* term in the equation 4.4 is computed by marginalization (Papoulis [1991]), i.e. by computing its integration over X as follows:

$$p(Y = y \mid M) = \int_{+\infty}^{-\infty} p(Y = y \mid X = x, M)p(X = x \mid M)dx. \tag{4.14}$$

The *evidence* term helps us to evaluate the quality of the assumed *prior* models, i.e. to access how good the assumed model fits to the given data. Considering a family of different models $(M_1, M_2, \cdots, M_n)$, the equation 4.14 will compute the likelihood of each model to facilitate the selection of most suitable model for the data y. It must be noted that its possible that more complex *prior* models will fit the data better. But at the same point Bayesian inference encapsulate Occam's razor (MacKay [1991]), which states that unnecessarily complex models are not preferred to simpler ones. It can also be said that a more plausible hypothesis seems to be simpler.

Bayesian image analysis makes use of explicit probability models to incorporate general and scene-specific prior knowledge into the processing of images and aims to provide a unified framework within which many different tasks can be considered (Besag [1989]).

4.3 Bayesian inference applied to SAR images

In the earlier part of this chapter we discussed the general framework of the first level of Bayesian inference for model-parameter estimation and the second level of Bayesian inference for model-selection. This section deals with the application of this to the information extraction focused on the SAR images. Our objective is to determine the model-parameters for different selected models. The idea of a Bayesian-based information extraction method is to find a *prior* model, which approximates the distribution of SAR image well and a *likelihood* model, which estimates the speckle within the SAR images, in order to obtain the *posterior* pdf. If we assume that the speckle is fully developed, the speckle corrupting the backscattered signal in the SAR image can modeled as multiplicative noise (Touzi [2002]), which can be represented as follows

$$y = x \cdot z, \tag{4.15}$$

where y represents an observed noisy SAR image, x denotes an ideal noise-free reflectance, and z stands for the speckle noise.

Now the Bayesian inference approach analogous to the multiplicative noise model for the SAR images can also be formulated. We can rewrite the Bayes' theorem shown in equation 4.4 in context to the SAR images as follows:

$$p(x \mid y, \boldsymbol{\theta}) = \frac{p(y \mid x)p(x \mid \boldsymbol{\theta})}{p(y \mid \boldsymbol{\theta})}, \tag{4.16}$$

where y represents the SAR image, x is the noise-free image and $\boldsymbol{\theta}$ stands for the unknown model parameters. The conditional pdf denoted by $p(x \mid y, \boldsymbol{\theta})$ indicates the *posterior* pdf, which stands for the ideal noise-free image, $p(y \mid x)$ represents the *likelihood* pdf, $p(x \mid \boldsymbol{\theta})$ denotes the *prior* pdf. The evidence term, $p(y \mid \boldsymbol{\theta})$, acts as a normalized constant in the first level of inference.

First level of Bayesian inference is used to obtain a model which fit the SAR data, estimating the maximum *a posteriori* (MAP) of the noiseless image. It is expressed by the *posterior* pdf and is obtained using the *likelihood* and *prior* pdfs together, as follows

$$\hat{x}(y) = \arg\max_{x} p(y \mid x)p(x \mid \boldsymbol{\theta}), \tag{4.17}$$

where $\hat{x}$ represents the de-speckled image. The second order Bayesian inference enables an estimation of parameter $\boldsymbol{\theta}$ using evidence maximization. This evidence characterizes how well the estimated data fits the original data. Selecting the best model, taking into account the speckle *likelihood*, the solution can be found by maximizing the *evidence* as a function of $\boldsymbol{\theta}$, as follows

$$p(y \mid \boldsymbol{\theta}) = \int p(y \mid x)p(x \mid \boldsymbol{\theta})dx, \tag{4.18}$$

While dealing with SAR images, *likelihood* and *prior* need to be defined accordingly. In the subsequent discussion, the *likelihood*, or the noise which is speckle in case of square root intensity SAR images is well-approximated by a Gamma distribution, written mathematically as follows:

$$p(y \mid x) = 2\left(\frac{y}{x}\right)^{2L-1} \frac{L^L}{x\Gamma(L)} \exp\left(-L\left(\frac{y}{x}\right)^2\right), \tag{4.19}$$

with L as the number of looks. For computation purposes, equivalent number of looks (ENL) is used defined as the number of independent intensity values averaged per pixel (Oliver & Quegan [1998]). The ENL over a homogeneous area can be expressed as: ENL = (Mean)^2/Variance.

The use of a Bayesian framework in image restoration and reconstruction, which we will apply for information extraction requires the specification of a *prior* distribution for the image (Besag [1974]). Now the critical issue is to construct this *prior* statistical models for SAR images, for which generally two approaches were followed. In the first one, images are modeled as Markov random fields (MRFs) (Chellappa & Jain [1993], Cross & Jain [1983]), and in the second, images are modeled as Gibbs random fields (GRFs) models (Besag [1974], Hassner & Sklansky [1980]). In both the MRFs and GRFs, a random field represents the ensemble of random variables in the form of a lattice to describe the statistical dependence and correlations between neighboring pixels. The image pixels in the SAR images (a 2-D signal) forms the basis for this lattice structure in the random fields.

In this thesis we restrict our discussion to two models relating to the GRFs family only: namely Gauss-Markov random fields (GMRF), and autobinomial Model (ABM). However, the Hammersley-Clifford Theorem (Hammersley & Clifford [1971]) asserts that for a limited neighborhood and discrete state of X, every Markov random field has an equivalent Gibbs random field. Before going into the theoretical details of GMRF and ABM *prior* pdfs along with the respective $\boldsymbol{\theta}$ parameters estimation methodology, we will briefly discuss the neighborhood systems and cliques.

4.3.1 Image description, neighborhood systems and cliques

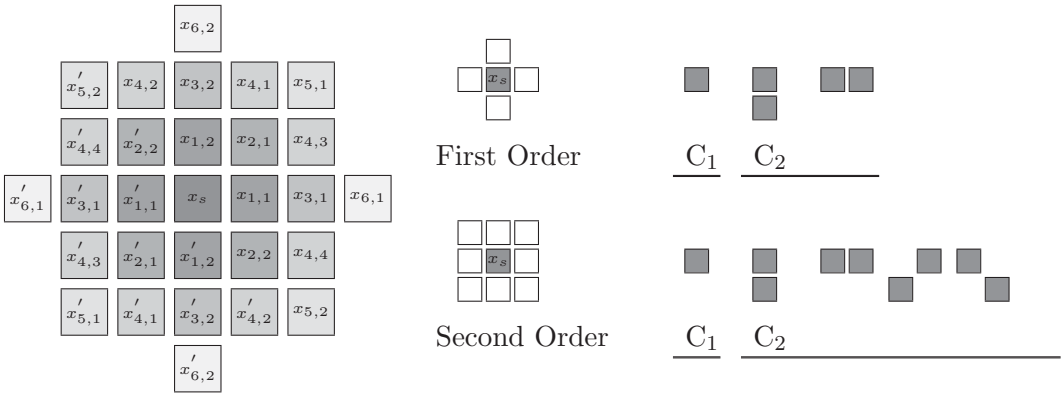
For a SAR image viewed as a 2-D stochastic process, the GRFs are defined on a rectangular lattice L of a finite size $M \times N$ with image pixels as points in the lattice. The neighborhood system, represented by Ω_s^m, around a given center pixel x_s is defined as (Li [1995]):

$$\Omega_s^m = \{t\} \quad \text{with} \quad \begin{cases} s \notin \Omega_s, \\ t \in \Omega_s, \quad \text{then} \quad s \in \Omega_t. \end{cases} \tag{4.20}$$

Here, $\{t\}$ is a set of all pixels in the vicinity of site s, excluding s. The hierarchical order of the sequence in image modeling follows a nearest neighborhood model, i.e. the closest four neighbors of each pixels occur in the sequence first. The maximum limit to the structural complexity of the GRF model is defined by the size of neighborhood system, i.e. the model-order. In the neighborhood system Ω_s^m, the model order is denoted by m. The neighborhood system in the lattice structure for model-orders 1 to 6, i.e. $\Omega_s^1, \Omega_s^1, \cdots, \Omega_s^6$ around a central pixel x_s are shown in the figure 4.1(left). The neighboring pixels around x_s are denoted as $x_{i,j}$, with i as the model-order. The index j is just a counting variable in neighborhood system. In the figure 4.1(center), first and second model-order neighborhood systems are specifically illustrated. It should be pointed out here that i and j are not related to the Cartesian coordinates in any sense.

The clique, denoted by c, is a subset of L associated with a neighborhood, which specifies the possible topology of the random structures in different realization of the stochastic process (Derin & Elliott [1987]). The clique c may consists of a single pixel or many pixels. Examples for single pixel clique and two-pixel clique for first and second model order are illustrated in the figure 4.1(right).

The length of the parameter vector $\bar{\theta}$ is determined by the model-order. The elements in $\bar{\theta}$ are represented as $\theta_{i,j}$, with i again as the model-order and j just a counting variable. It can be seen from the figure 4.1(left), that the length of parameter vector $\bar{\theta}$ for the model-order 4 will be 10, and for the model-order 6 will be 14.



$$\bar{\theta} = [\theta_{11},\theta_{12},\theta_{21},\theta_{22},\theta_{31},\theta_{32},\theta_{41},\theta_{42},\theta_{43},\theta_{44},\theta_{51},\theta_{52},\theta_{61},\theta_{62}]$$

Figure 4.1: Neighborhood system and clique. (Left) Illustration of neighborhood system Ω_s^m with model order $m = 6$ and corresponding model parameter vector $\bar{\theta}$. For a given center pixel x_s, the neighborhood pixels at order k are denoted with $x_{k,l}$. (Center) First and second order neighborhood for a given center pixel x_s. (Right) One-pixel clique - C_1 and two-pixels clique - C_2 configurations for corresponding first and second model order

4.3.2 Energy and potential functions

Once the neighborhood system and clique for a certain model-order are assigned, we need to define the equivalent assumed GRF on this lattice. The joint distribution of the GRF models are given in the form of an exponential distribution as follows (Derin & Elliott [1987]):

$$p(X = x) = \frac{1}{Z} \mathrm{e}^{-U(x)}. \tag{4.21}$$

Here, $U(x)$ is the the energy function and $Z = \sum_x \mathrm{e}^{-U(x)}$ is the partition function, a normalizing constant. Frequently, the exponent in the equation 4.21 is expressed as $(1/T)U'(x)$, with T as the temperature - a scale hyper-parameter, and $U'(x)$ as the energy function. The energy function is represented in terms of the sum of potential functions $V_c(x)$ associated with clique c, as follows:

$$U(x) = \sum_{\text{all cliques}} V_{\text{clique}}(x, \bar{\theta}), \quad \text{with } \bar{\theta} \text{ as parameter vector.} \tag{4.22}$$

The potential function provides the parametric description of the type of interactions between the samples of the respective random field inside the clique structure. The only condition on the otherwise totally arbitrary clique potential $V_{\text{clique}}(x, \bar{\theta})$ is that it depends only on the pixel values in clique c (Derin & Elliott [1987]).

4.4 GRFs model parameters for information extraction in SAR images

In literature, many GRFs models can be found (Li [1995]). As mentioned earlier, in this thesis we will specifically discuss only two models: GMRF and ABM. A MAP estimate applied to GRFs using evidence maximization was proposed initially as a framework for model-based despeckling (MBD) and information extraction in SAR images in Datcu *et al.* [1998], and Walessa & Datcu [2000]. An extensive study of MBD performance of the GMRF and ABM was carried out in Molina [2011]. However, our objective in this thesis is the implementation and evaluation of the GRFs model for information extraction in SAR images for two specific applications: pixel-based classification and patch-oriented image categorization. Adaptation of the GMRF to the complex-valued SAR images presented in Singh *et al.* [2009], and Soccorsi [2010] will also be discussed.

The parameter estimation using MAP, where the maximum of the product of *prior* and *likelihood* is estimated by optimizing the values of the model-parameters $\boldsymbol{\theta}$. This can be achieved using expectation maximization, analytical parameter estimation, or iterative evidence maximization (IEM) etc. For our experiment, we will be using the IEM method for the parameter estimation for GMRF and ABM *prior* pdfs. The IEM was proposed in Datcu *et al.* [1998], Walessa & Datcu [2000] by means of a GMRF model as *prior*, which was later adapted to the ABM as *prior* in Hebar *et al.* [2009]. Before going into the details of $\text{MAP}_{\text{IEM}} - \text{GMRF}$, $\text{MAP}_{\text{IEM}} - \text{ABM}$, and adaptation of GMRF to the complex-valued SAR images, we will briefly discuss the expectation maximization and iterative evidence maximization framework in context to the MAP parameter estimation.

- **Expectation maximization**

If we can formulate the solution for the MAP estimate using the first level of Bayesian inference (refer the equation 4.17), which is the estimate of the noise-free SAR image pixel $\hat{x}$, the problem remains how to choose the values of the model parameter vector $\boldsymbol{\theta}$. One popular method in statistics is the expectation maximization (EM) method. The EM method is an iterative method, where first, a function for the expectation of the log-likelihood is created (Dempster *et al.* [1977]). This function is defined as follows:

$$Q(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}^c) = \int \log\Big(p(\hat{\boldsymbol{\theta}} \mid X = x, Y = y)\Big)\, p(X = x \mid Y = y, \hat{\boldsymbol{\theta}}^c) dx. \tag{4.23}$$

This integral is very complex and very challenging to have an analytical solution for the complex models. In practice, an EM-type method is used with a simplification of the function Q. The method consists of two steps: (E)xpectation-Step and (M)aximization-Step. In the E-Step, MAP estimate $\hat{x}^{\text{MAP}}$ is computed with current estimate of model parameters $\hat{\boldsymbol{\theta}}^c$. In the M-Step, using this MAP estimate $\hat{x}^{\text{MAP}}$, ML estimate of $\boldsymbol{\theta}$ is computed by maximizing $p(X = \hat{x}^{\text{MAP}})$. The process is iterated until convergence is achieved.

- **Iterative evidence maximization**

To solve the problem of suitable $\boldsymbol{\theta}$ parameter estimation that best explains the noisy image $\mathbf{y}$ for a given speckle noise level L, another method taking advantage of the second level of Bayesian inference was proposed in Datcu *et al.* [1998]. This method, called the iterative evidence maximization (IEM), works on the principle of maximization of an approximation of the *evidence*

$p(Y = y \mid \boldsymbol{\theta}, M)$, with M as the selected *prior* pdf. The general IEM framework is shown in the figure 4.2.

The objective is to search for suitable $\boldsymbol{\theta}$ parameters to have a good estimate of $\hat{x}$. The IEM framework is initialized with an initial guess of $\boldsymbol{\theta}$ parameters, with $\theta^i = k$ for all i, where k is equal to 0.5/ {size of $\boldsymbol{\theta}$ vector}. In (I)nitial-step, using these initial values of $\boldsymbol{\theta}$ parameters, a first MAP estimate of $\hat{x}$ is obtained using the first level of Bayesian inference. With the values of $\boldsymbol{\theta}$ and $\hat{x}$, *evidence* is computed using the second level of Bayesian inference. In the (E)vidence (M)aximization-Step, keeping $\hat{x}$ fixed, the value of $\boldsymbol{\theta}$ parameters is modified iteratively in order to maximize the approximation of the evidence. The EM-Step is repeated until convergence is reached.

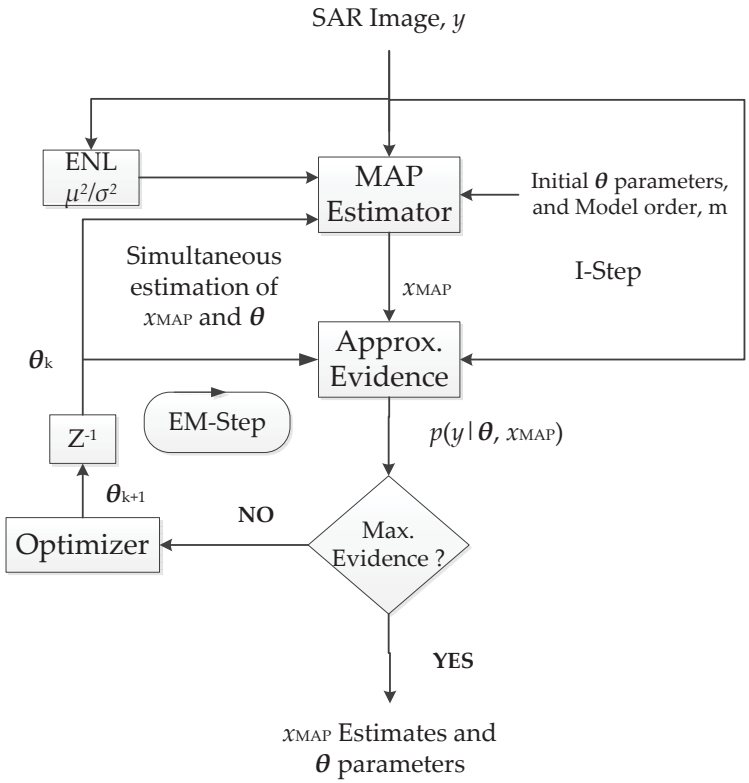


Figure 4.2: Iterative evidence maximization (IEM) framework for maximum *a posteriori* estimation. The I-step stands for initial step, and EM-step here stands for evidence maximization, where an approximate evidence term is used.

4.4.1 Gauss-Markov random fields model

GMRF was first introduced in Kashyap & Chellappa [1983] in the family of conditional Markov (CM) models. This model is a rather simple image model with the energy function presented in the equation 4.22 as:

$$U(x_s \mid \Omega_s, \boldsymbol{\theta}) = \frac{(x_s - \eta_s)^2}{2\sigma^2}, \tag{4.24}$$

With, $\boldsymbol{\theta}$, σ are the model parameters. The interaction of pixels in an image with its neighbors is explained by the neighborhood system around the central pixel, η_s, written as follows:

$$\eta_s = \sum_{ij} \theta_{i,j} \cdot (x_{ij} + x'_{ij}) \tag{4.25}$$

The corresponding pdf of the stochastic GMRF model, following a Gaussian auto-regressive (AR) model (Schistad Solberg & Jain [1997]), can be expressed as follows:

$$p(x_s \mid \boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_s - \eta_s)^2}{2\sigma^2}\right). \tag{4.26}$$

The neighborhood system η_s will be defined by the model order m as demonstrated in the section 4.3.1, which will also control the length of model parameter vector $\boldsymbol{\theta}$.

4.4.1.1 MAP estimation and iterative evidence maximization for GMRF

At the first level of Bayesian inference, the MAP estimates of the noise-free pixels in SAR images can be obtained by finding the maximum of the product of the *likelihood* and *prior*, while ignoring the *evidence* at this stage, as follows:

$$p(x_s \mid y_s) = p(y_s \mid x_s, \boldsymbol{\theta}) p(x_s \mid \boldsymbol{\theta}). \tag{4.27}$$

Our objective is to maximize this quantity, which is achieved by setting the first derivative of $p(x_s \mid y_s)$ to zero. For simplicity, the derivative of the logarithm of $p(X = x \mid Y = y)$ is used. This will not change the location of the maxima as it is a monotonic function. Assuming the *likelihood* to follow the Gamma distribution from the equation 4.19, and the *prior* for GMRF from the equation 4.26, we obtain the MAP equation as follows:

$$\begin{aligned} \frac{\partial}{\partial x_s} \log p(x_s \mid y_s) = \frac{\partial}{\partial x_s} \log \left[2 \left(\frac{y_s}{x_s}\right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \mathrm{e}^{-L\left(\frac{y_s}{x_s}\right)^2} \right] \\ + \frac{\partial}{\partial x_s} \log \left[\frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x_s - \mu_s)^2}{2\sigma^2}} \right]. \end{aligned} \tag{4.28}$$

Setting the left-hand side in the equation 4.28 to zero, a fourth-order polynomial is obtained with four possible roots, as follows:

$$x_s^4 - \mu_s x_s^3 + 2L\sigma^2 x_s^2 - 2L\sigma^2 y_s^2 = 0. \tag{4.29}$$

Since the valid x_{MAP} estimate, i.e. x_s should be real-valued and positive, convergence of iterative conditional modes (ICM) algorithms for relaxation can be employed to obtain the desired solution for the maximum of the *a posterior* (Metropolis *et al.* [1953]). A few iterations of a simple steepest descent algorithm with as initial guess provided good results (Walessa & Datcu [2000]).

The above x_{MAP} in the I-Step shown in the figure 4.2, is obtained using some initial values of $\boldsymbol{\theta}$ parameters. The next issue is to obtain the values of $\boldsymbol{\theta}$ parameters, which should explain the SAR image better. This is achieved by using the second level of Bayesian inference, shown as EM-Step in the figure 4.2 by IEM. Recalling the formulation of the *evidence* term from the equation 4.18:

$$p(y_s \mid \boldsymbol{\theta}) = \int p(y_s \mid x_s, \boldsymbol{\theta}) p(x_s \mid \boldsymbol{\theta}) dx. \tag{4.30}$$

The closed-form analytical solution of this integral can not be obtained. To make the problem of computation of the *evidence* term tractable, several approximations have been proposed in Walessa & Datcu [2000]. The integrand in the equation 4.30 is assumed to be a product of mutually-independent random variables, thus its joint pdf can be expressed as the product of pdfs of its components. Of course, we only assume the statistical independence which is a good approximation when the size $N \times N$ of image under consideration is large enough. Now the multidimensional pdf is approximated by a multivariate Gaussian pdf with a Hessian matrix ($\mathbf{H}$, which is centered around the MAP estimate of x, i.e., around the maximum of the *posterior* distribution (MacKay [1991], Sivia [1996]). Consequently, the Gaussian-shaped multivariate *posterior* gives the approximation of the evidence as follows:

$$p(y_s \mid \boldsymbol{\theta}) = \int p(y_s \mid x_s, \boldsymbol{\theta}) p(x_s \mid \boldsymbol{\theta}) dx, \tag{4.31}$$

$$\approx \int \prod_{s=1}^{N \times N} p(y_s \mid x_s^{MAP}) p(x_s^{MAP} \mid \theta) \cdot \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^{MAP})^T} \mathbf{H}(\mathbf{x}-\mathbf{x}^{MAP}) dx, \tag{4.32}$$

$$\approx \frac{(2\pi)^{\frac{|N \times N|}{2}}}{\sqrt{\det \mathbf{H}}} \prod_{s=1}^{|N \times N|} p(y_s \mid x_s^{MAP}) p(x_s^{MAP} \mid \theta). \tag{4.33}$$

Here, $(N \times N)$ represents whole set of pixels of the SAR image with its cardinal $\mid N \times N \mid$, and x_s^{MAP} is the MAP estimate of x_s obtained using the fixed parameter vector $\boldsymbol{\theta}$ obtained in M-Step in the figure 4.2, computed by following the equation 4.29. The Hessian matrix $\mathbf{H}$ in the equation 4.33 is the a square matrix of the second-order partial derivatives of an uni-variate function written as follows:

$$\mathbf{H} = -\nabla\nabla \sum_{s=1}^{N \times N} \log(p(y_s \mid x_s^{\mathrm{MAP}}) p(x_s^{\mathrm{MAP}} \mid \boldsymbol{\theta})). \tag{4.34}$$

The determinant of the Hessian matrix $\mathbf{H}$, with h_{ss} as its diagonal elements, can be approximated as follows:

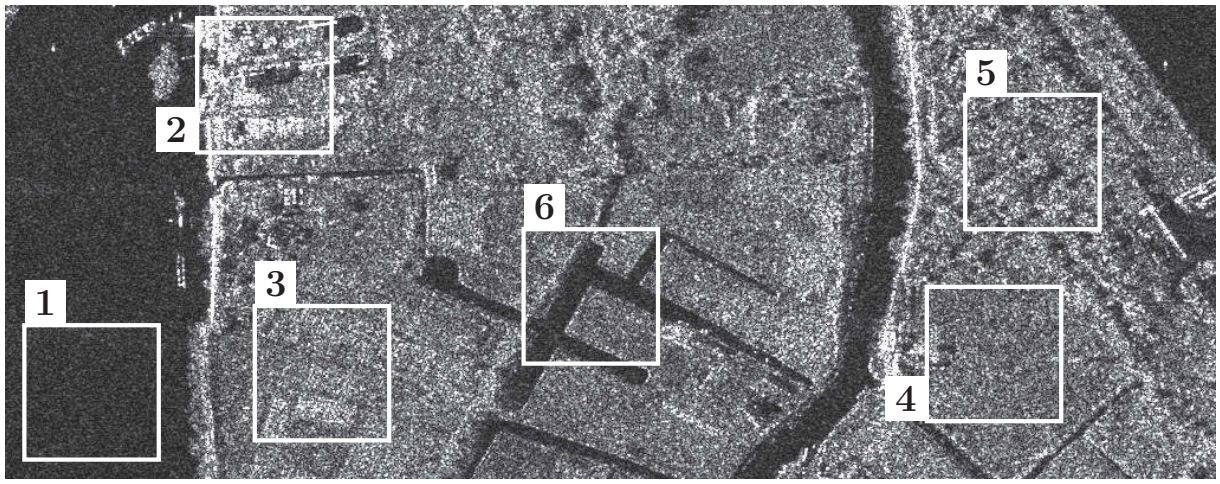
$$\det \mathbf{H} \approx \prod_{s=1}^{N \times N} h_{ss}. \tag{4.35}$$

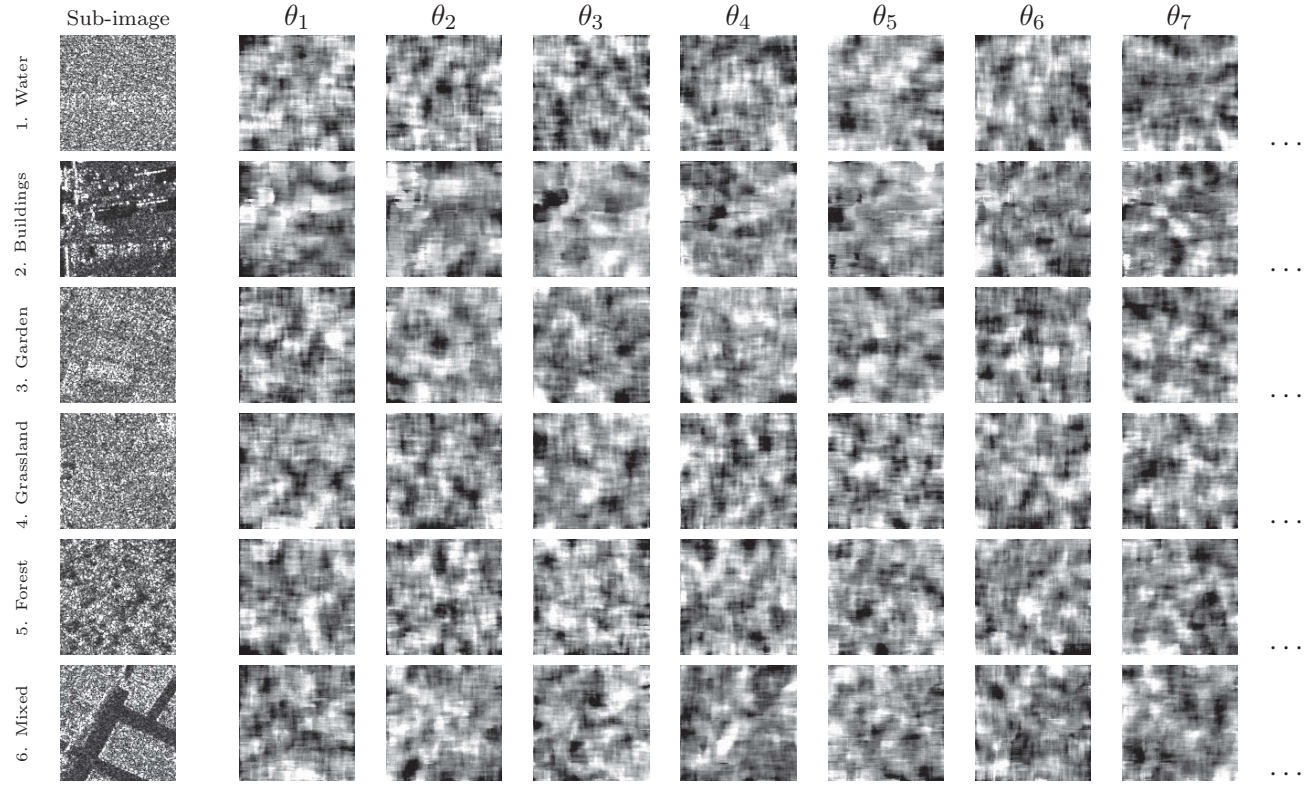
Another approximation for the diagonal elements of Hessian matrix can be written as follows:

$$h_{ss} = \frac{6Ly_s^2}{x_s^{4\mathrm{MAP}}} - \frac{2L}{x_s^{2\mathrm{MAP}}} + \frac{1}{\sigma^2} \left(1 + \sum_{k \in \Omega_s} \theta_k^2 \right) \tag{4.36}$$

The above approximations in the evidence maximization step help us to obtain a feasible solution for $\hat{x}$ and $\boldsymbol{\theta}$ by avoiding a highly computational demanding Markov Chain Monte Carlo (MCMC) method (von der Linden *et al.* [1999]). Starting with a set of initial $\boldsymbol{\theta}$ parameter, using the IEM framework shown in the figure 4.2, will provide us the final estimate for noise free pixels $\hat{x}$, and the final $\boldsymbol{\theta}$ parameter vector.

This $\boldsymbol{\theta}_{\mathrm{GMRF}}$ parameter vector will be later used as a feature descriptor for classification or categorization purposes assuming that it explains the texture and image content, provided our assumptions of GMRF as *prior* pdf are plausible. An example of estimated $\boldsymbol{\theta}_{\mathrm{GMRF}}$ parameters using MAP following IEM is presented in the figure 4.3.





Sub-image	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	$\cdots$
1.	0.2021	0.0067	-0.0333	0.0544	-0.0024	-0.0042	0.1135	0.0044	-0.0788	-0.0092	$\cdots$
2.	0.2576	-0.0115	-0.0229	-0.0492	-0.0279	0.0258	0.1948	0.1162	0.0012	-0.0252	$\cdots$
3.	0.1935	-0.0056	-0.0183	0.0930	0.0017	-0.0052	0.2821	-0.0252	0.0085	-0.0702	$\cdots$
4.	0.2912	0.0003	-0.0083	0.0885	0.0239	-0.0383	0.3030	-0.0220	-0.0733	-0.0121	$\cdots$
5.	0.3062	0.0421	-0.0974	0.0635	0.0162	0.0630	0.4539	0.0712	-0.0488	-0.0829	$\cdots$
6.	0.2948	-0.0447	-0.0492	0.0553	-0.0256	0.0376	0.3258	0.0421	-0.0629	-0.0233	$\cdots$

Figure 4.3: Parameter estimation for GMRF as *prior* pdf for various textures with analyzing window of size 16×16 pixels and model order=5. The values of texture parameters in the table are the θ values for the center pixel in each sub-image. The variation in θ value shows that it captures the characteristic spatial structures. The higher the model order is, the better is the texture discrimination that can be achieved. The parameter computation using IEM is computationally very expensive. It took around 10 minutes to process one sub-image at presented settings on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM.

4.4.2 Autobinomial Model

The ABM model was originally presented in Cross & Jain [1983]. The use of ABM model for remote-sensing images was investigated in Schroder *et al.* [1998]. The ABM is able to generate blurry, sharp, line-like and blob-like textures which make the ABM generated texture more natural than the GMRF. The assumption that the ABM would fit the SAR data better than GMRF, and the ABM included into the MAP estimator would be able to estimate more textures than GMRF, has been postulated in Hebar *et al.* [2009]. The energy function in ABM model is defined as:

$$U(x_s \mid \Omega_s, \boldsymbol{\theta}) = -\log \binom{G}{x_s} - x_s \cdot \eta_s. \tag{4.37}$$

Here, G is the maximum gray value in the analyzed image, with x_s as the observed pixel. The binomial coefficients are denoted as $\binom{n}{m}$. The term η_s, denoting the effective influence of neighboring pixels is defined as follows:

$$\eta_s = a + \sum_{ij} b_{ij} \frac{(x_{ij} + x_{ij}^{'})}{G} \tag{4.38}$$

The x_{ij} and $x_{ij}^{'}$ are the neighboring pixels as shown in the figure 4.1. The corresponding pdf in the equation 4.21 for the stochastic ABM is given as follows:

$$p(x \mid \boldsymbol{\theta}) = \binom{G}{x_s} \rho_s^{x_s} (1-\rho_s)^{G-x_s}, \tag{4.39}$$

Here, again the $\boldsymbol{\theta}$ is the parameter vector, and ρ_s is the function of model parameter vector $\boldsymbol{\theta} = (a, b_{ij})$, written as $\rho_s = 1 + \exp(-\eta_s)^{-1}$.

4.4.2.1 MAP estimation and iterative evidence maximization for ABM

At the first level of Bayesian inference, the MAP estimates of the noise-free pixels in SAR images can be obtained by the finding the maximum of the product of the *likelihood* and *prior*, i.e. setting the first derivative of $p(x_s \mid y_s) = p(y_s \mid x_s, \boldsymbol{\theta}) p(x_s \mid \boldsymbol{\theta})$ to zero. Again, for simplicity, the derivative of the logarithm of $p(x_s \mid y_S)$ is used. The *likelihood* is again assumed to be Gamma distribution as per the equation 4.19. We obtain our MAP equation for ABM by introducing differentials of the logarithm as follows:

$$\begin{aligned} \frac{\Delta}{\Delta x_s} \log p(x_s \mid y_s) = \frac{\Delta}{\Delta x_s} \log \left[2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \mathrm{e}^{-L\left(\frac{y_s}{x_s}\right)^2} \right] \\ + \frac{\Delta}{\Delta x_s} \log \left[\binom{G}{x_s} \rho_s^{x_s} (1-\rho_s)^{G-x_s} \right]. \end{aligned} \tag{4.40}$$

Setting the left-hand side in the equation 4.40 and solving the differentials will provide us the x^{MAP} estimates. In the equation 4.40, the analytical solution for the first term exists, but the second term needs to be solved numerically by subtracting $\log p((x_s + 1 \mid \theta_s)) - \log p(x_s \mid \theta_s)$. The result is given by the following simplified formulation (Hebar *et al.* [2009]):

$$-\frac{2L}{x_s} + 2L\frac{y_s^2}{x_s^3} + \log\left(\frac{G - x_s}{x_s + 1}\right) + \log\left(\frac{\rho_s}{1-\rho_s}\right). \tag{4.41}$$

At the second level of Bayesian inference, we need to solve the integral $p(y_s \mid \boldsymbol{\theta}) = \int p(y_s \mid x_s, \boldsymbol{\theta}) p(x_s \mid \boldsymbol{\theta}) dx$ to compute the *evidence*. As the integrand is again very complex, similar

to the GMRF case, we need to do several approximations. First we break apart the joint pdf and express it in terms of the product of pdfs of its components, assuming there statistical independence. Now again, the multidimensional pdf is approximated by a multivariate Gaussian pdf with a Hessian matrix $\mathbf{H}$, which is centered around the MAP estimate of x, i.e., around the maximum of the posterior distribution. These approximations can be mathematically written as follows:

$$p(y_s \mid \boldsymbol{\theta}) = \int p(y_s \mid x_s, \boldsymbol{\theta}) p(x_s \mid \boldsymbol{\theta}) dx, \tag{4.42}$$

$$\approx \int \prod_{s=1}^{N \times N} p(y_s \mid x_s) p(x_s^{\text{MAP}} \mid \boldsymbol{\theta}) \exp(-0.5 \Delta x_s^{\text{T}} \mathbf{H} \Delta x_s) \tag{4.43}$$

$$\approx \frac{2\pi^{(N \times N)/2}}{\sqrt{\mid \mathbf{H} \mid}} \prod_{s=1}^{N \times N} p(y_s \mid x_s^{\text{MAP}}) p(x_s^{\text{MAP}} \mid \boldsymbol{\theta}) \tag{4.44}$$

The Hessian matrix $\mathbf{H}$ is again the second order partial derivative of a uni-variate function, written same as the equation 4.34:

$$\mathbf{H} = -\nabla\nabla \sum_{s=1}^{N \times N} \log(p(y_s \mid x_s^{\text{MAP}}) p(x_s^{\text{MAP}} \mid \boldsymbol{\theta})). \tag{4.45}$$

Another simplification is to use the logarithm form, to obtain:

$$\begin{aligned} \log p(y \mid \boldsymbol{\theta}) = \sum_{s=1}^{N \times N} \Bigg[\frac{(N \times N) \log 2\pi - \log h_{ss}}{2} + \log \left(2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \mathrm{e}^{-L\left(\frac{y_s}{x_s}\right)^2} \right) \\ + \log \left(\binom{G}{x_s} \rho_s^{x_s} (1 - \rho_s)^{G - x_s} \right) \Bigg], \end{aligned} \tag{4.46}$$

where, h_{ss} are the main diagonal elements of the Hessian matrix $\mathbf{H}$, approximated as follows:

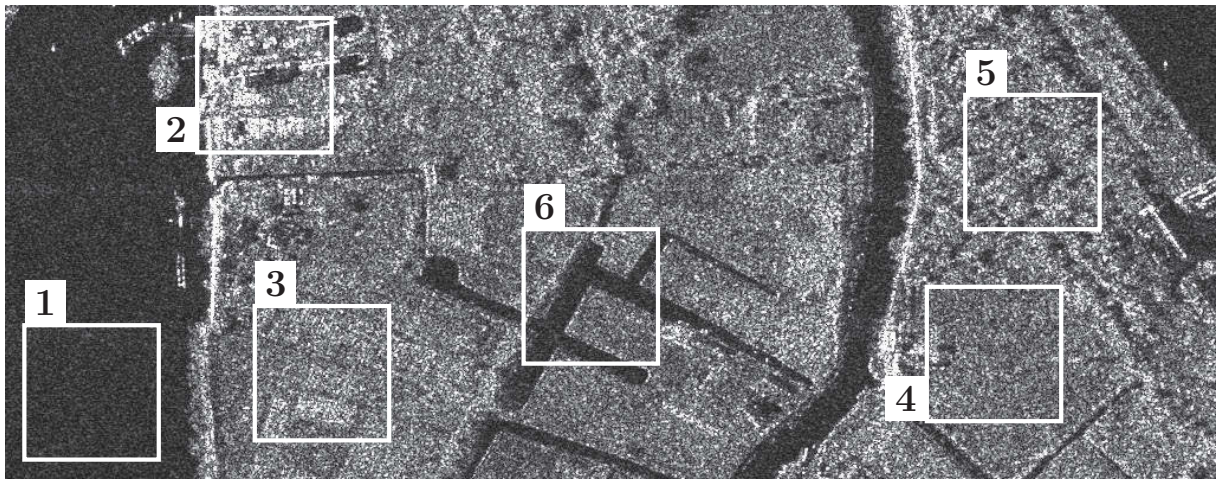
$$h_{ss} \approx \sum_{s=1}^{N \times N} \left(\frac{2L}{x_s^{\text{MAP}\,2}} - 6L \frac{y_s^2}{x_s^{\text{MAP}\,4}} - \frac{1}{G - x_s^{\text{MAP}}} - \frac{1}{x_s^{\text{MAP}} + 1} \right). \tag{4.47}$$

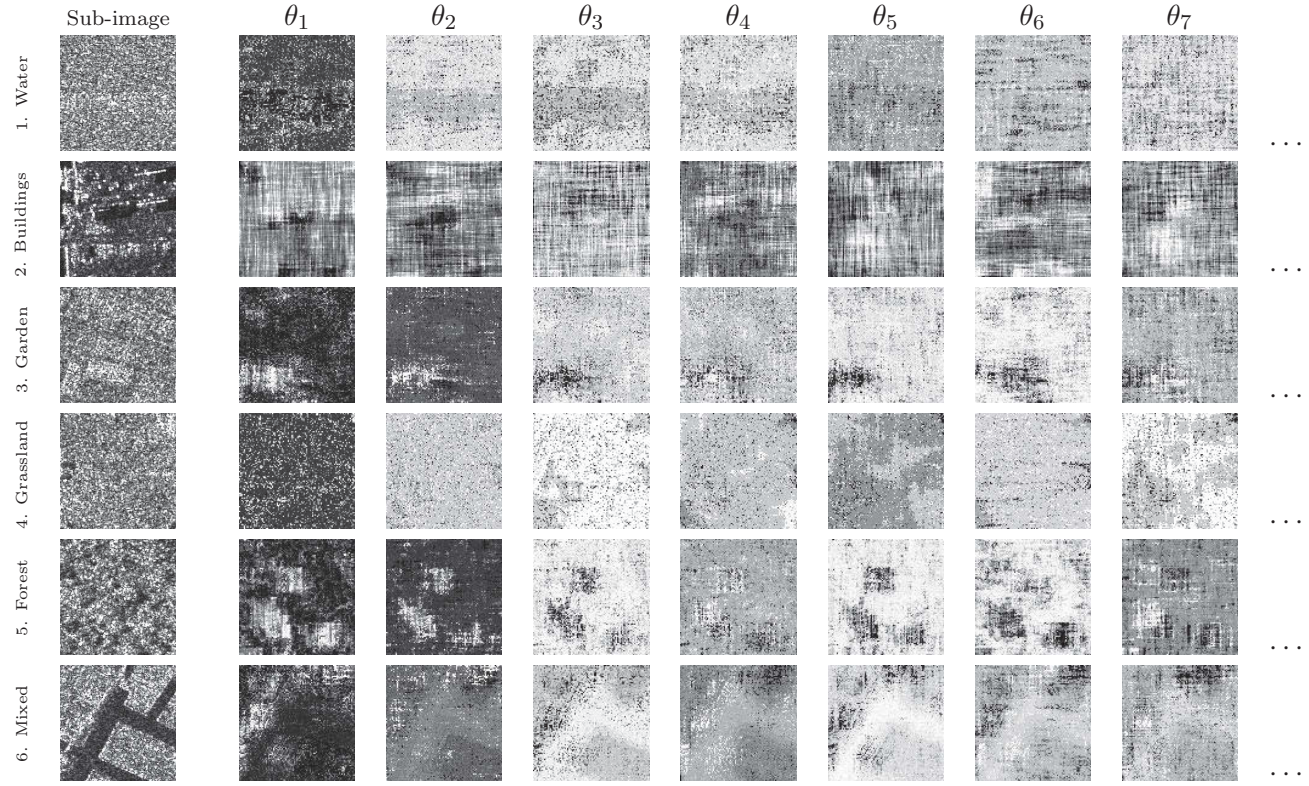
Again, we use only the diagonal elements to obtain the determinant of the Hessian matrix $\mathbf{H}$ as follows:

$$\det \mathbf{H} \approx \prod_{s=1}^{N \times N} h_{ss}. \tag{4.48}$$

With these approximations, we will obtain the final estimate for noise free pixels $\hat{x}$, and the final $\boldsymbol{\theta}$ parameter vector after starting with an initial set of $\boldsymbol{\theta}$ parameters using the IEM framework shown in the figure 4.2.

This $\boldsymbol{\theta}_{\text{ABM}}$ parameter vector will be later used as a feature descriptor for classification or categorization purposes. An example of estimated $\boldsymbol{\theta}_{\text{ABM}}$ parameters using MAP following IEM is presented in the figure 4.4.





Sub-image	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	$\cdots$
1.	0.2467	0.2367	0.1167	0.1867	0.1667	0.2267	0.1167	0.2067	0.1667	0.1467	$\cdots$
2.	0.4067	0.2567	0.0967	0.1267	0.1067	0.1467	0.1567	0.1667	0.1467	0.1867	$\cdots$
3.	0.2067	0.2067	0.1867	0.1867	0.1867	0.1467	0.2067	0.1767	0.1367	0.1867	$\cdots$
4.	0.2267	0.2067	0.1767	0.2067	0.1967	0.2167	0.1567	0.1767	0.1967	0.1767	$\cdots$
5.	0.2267	0.1867	0.1867	0.1767	0.1767	0.1767	0.1867	0.1767	0.1367	0.2067	$\cdots$
6.	0.2067	0.2067	0.2067	0.2067	0.1967	0.1967	0.1867	0.1867	0.1767	0.1767	$\cdots$

Figure 4.4: Model parameters for ABM model as *prior* pdf with analyzing window size of 16×16 pixels and model order=5. The values of texture parameters in the table are the θ values for center pixel in each sub-image. The variation in θ value shows that it captures the characteristic spatial structures. In comparison to the GMRF as *prior* pdf, ABM seems to capture the textural information better. E.g. in sub-image 3, the variation in the garden texture at lower left corner is captured. As well as in sub-image 5, the collection of trees in a forest forming detailed texture is also captured. But the parameter computation for ABM using IEM is even more computationally expensive than GMRF. It took around 60 minutes to process one sub-image, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM at the presented settings.

4.4.3 Gauss-Markov random fields model for complex-valued SAR images

A complex random process with its realization as a complex-valued random variable $\mathbf{x} = x_R + jx_I$ can be represented through a pair of real vectors, with real x_R and imaginary x_I parts as its components. Thus it is also possible to extend the GMRF model presented in section 4.4.1 for amplitude SAR images to the complex-valued SAR images as proposed in Soccorsi [2010]. Let us recall from the equation 4.24, the energy function for the GMRF model:

$$U(x_s \mid \Omega_s, \boldsymbol{\theta}) = \frac{(x_s - \sum_{ij} \theta_{i,j} \cdot (\mathbf{x}_{ij} + \mathbf{x}'_{ij}))^2}{2\sigma^2}. \tag{4.49}$$

The corresponding pdf of the GMRF model, following a Gaussian AR model, can be expressed as follows:

$$p(x_s \mid \boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_s - \eta_s)^2}{2\sigma^2}\right). \tag{4.50}$$

The associated complex auto-regressive process will have the form:

$$\mathbf{x}_s = \frac{1}{2} \sum_{ij} \boldsymbol{\theta}(\mathbf{x}_s + \mathbf{x}_s{}') + \mathbf{e}_s, \tag{4.51}$$

with $\mathbf{e}_s = e_{sR} + je_{sI}$ as the realization of the independent and identically distributed (iid) Gaussian process (Picinbono & Bouvet [1984]). We represent the complex-valued SAR pixel as $\mathbf{x}_s = x_{sR} + jx_{sI}$, similarly the model-parameters are $\boldsymbol{\theta} = \theta_R + j\theta_I$. Because of the isomorphism between real and imaginary planes $\mathbb{R}^2$ in the complex-valued space $\mathbb{C}$, the formulation for $\mathbf{x}_s$ can be vectorially written as follows:

$$\begin{bmatrix} x_{sR} \\ x_{sI} \end{bmatrix} = \frac{1}{2} \sum_{ij} \begin{bmatrix} \theta_R & -\theta_I \\ \theta_I & \theta_R \end{bmatrix} \begin{bmatrix} x_{ijR} + x'_{ijR} \\ x_{ijI} + x'_{ijI} \end{bmatrix} + \begin{bmatrix} e_{sR} \\ e_{sI} \end{bmatrix} \tag{4.52}$$

This allows us to represent the *prior* pdf of the GMRF model adapted to the complex-valued SAR data as follows:

$$p(\mathbf{x}_s \mid \boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\begin{bmatrix} x_{sR} \\ x_{sI} \end{bmatrix} - \boldsymbol{\eta}_s\right)^2}{2\sigma^2}\right), \quad \text{with } \boldsymbol{\eta_s} = \sum_{ij} \begin{bmatrix} \theta_R & -\theta_I \\ \theta_I & \theta_R \end{bmatrix} \begin{bmatrix} x_{ijR} + x'_{ijR} \\ x_{ijI} + x'_{ijI} \end{bmatrix}. \tag{4.53}$$

With these formulations, we can compute model parameters for real and imaginary parts of the complex-valued SAR images separately.

4.4.3.1 Model parameter estimation following a linear model approach

In case of the complex-valued SAR images, because of computational constrains, we are not going to apply the IEM framework as we did for the amplitude SAR image in section 4.4.1. Due to symmetry and linearity of the model and exploiting the matrix formulation, we can express the image pixels in accordance to the linear model from the equation 4.51 as follows:

$$x_s = \mathbf{G}\boldsymbol{\theta} + e_s, \tag{4.54}$$

with, x_s as the real part (x_{sR}) or imaginary part (x_{sI}) of the complex-valued pixel $\mathbf{x} = x_{sR} + jx_{sR}$, the matrix $\mathbf{G}$, a linear transformation, is matrix of cliques constructed from the image

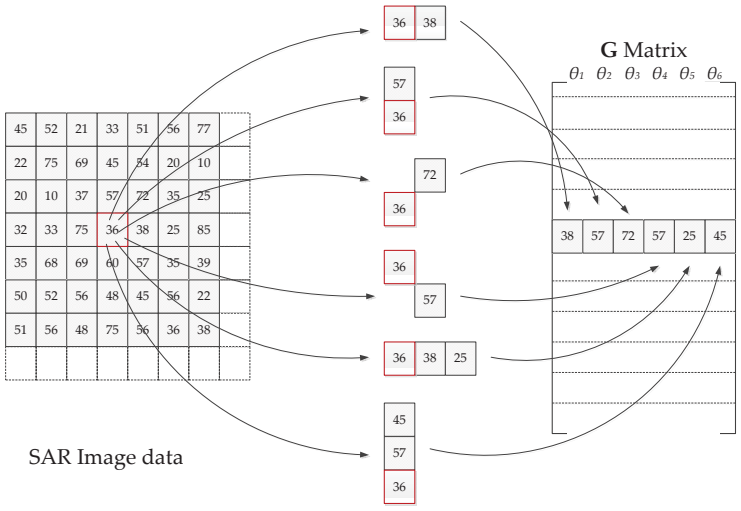


Figure 4.5: Architecture of formation of $\mathbf{G}$ matrix of cliques for the least square estimation of GMRF model-parameter vector $\boldsymbol{\theta}$.

data as shown in figure 4.5 (Datcu *et al.* [2004]), and e_s is again the realization of the iid Gaussian process. Thus, the unknown parameter vector $\boldsymbol{\theta}$ is computed using the least-square approximation methods (Ruanaidh & Fitzgerald [1996]) as follows:

$$\hat{\boldsymbol{\theta}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T x_s. \tag{4.55}$$

With, $(\mathbf{G}^T\mathbf{G})^{-1}$ as the pseudo-inverse matrix. The number of columns in $\mathbf{G}$ is equal to the size of model parameter vector $\boldsymbol{\theta}$, and number of rows is equal to the number of element in x_s. Here the data x_s forms a column vector consisting of the image pixels organized in a lexicographic order. The model variance can be computed as follows:

$$\sigma^2 = \mathbb{E}\left[(x_s - \mathbf{G}\hat{\boldsymbol{\theta}})^2\right]. \tag{4.56}$$

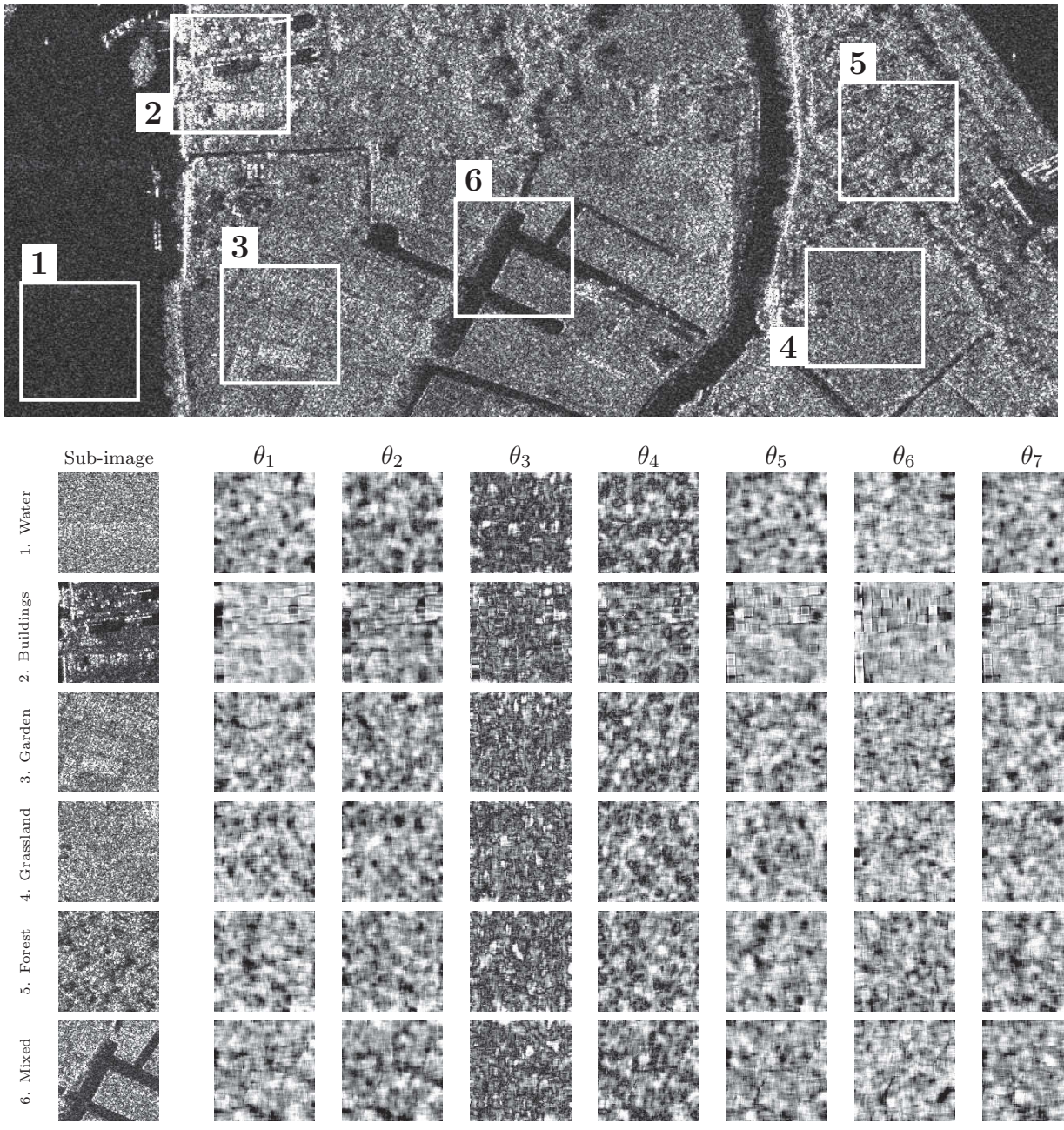
The model variance is the expectation of the difference between the data x_s and the best fit $\mathbf{G}\hat{\boldsymbol{\theta}}$ of the model over the data, $\hat{\boldsymbol{\theta}}$ is the complex-valued model parameter estimates. The information on the goodness of the model fit to the data is given by the model evidence, which can be written as follows (Ruanaidh & Fitzgerald [1996]):

$$p(x_s \mid \mathbf{G}, e) \approx \frac{\pi^{-N/2}\Gamma(q/2)\Gamma((n-q)/2)\det(\mathbf{G}^T\mathbf{G})^{-1})}{\kappa\boldsymbol{\theta}^T\boldsymbol{\theta}^{q/2}\hat{\sigma}^{n-q}}. \tag{4.57}$$

Here, n is the number of image pixels, q is the number of parameters in model parameter vector $\boldsymbol{\theta}$, κ is a normalization constant, and $\Gamma(\cdot)$ is the Gamma function.

However, along with $\boldsymbol{\theta}_{\text{C-GMRF}}$ parameters, the model variance σ^2 from the equation 4.56 and model evidence $p(x_s \mid \mathbf{G}, e)$ from the equation 4.57 can also be utilized while forming a feature-descriptor. We will be using only the model parameter vector $\boldsymbol{\theta}$, obtained using the least-square estimation approach, with a linear model framework to generate a feature-descriptor for the SAR image, renaming it as $\boldsymbol{\theta}_{\text{C-GMRF}}$.

This $\boldsymbol{\theta}_{\text{C-GMRF}}$ parameter vector will be later used as a feature descriptor for classification or categorization purposes. An example of estimated $\boldsymbol{\theta}_{\text{C-GMRF}}$ parameters is presented in the figure 4.6.



Sub-image	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	$\cdots$
1.	0.0250	0.0259	0.1140	0.0219	0.0603	0.0223	0.0334	0.0174	0.0246	0.0226	$\cdots$
2.	0.0255	0.0297	0.0548	0.0146	0.0441	0.1366	0.0630	0.0751	0.0573	0.0355	$\cdots$
3.	0.0575	0.0154	0.0582	0.0535	0.0476	0.1012	0.0925	0.0879	0.1107	0.0781	$\cdots$
4.	0.0643	0.0373	0.0443	0.0529	0.0839	0.1048	0.0591	0.0388	0.0637	0.0913	$\cdots$
5.	0.0861	0.0790	0.0811	0.0445	0.0970	0.0765	0.1013	0.0785	0.0495	0.1470	$\cdots$
6.	0.0115	0.0932	0.0845	0.1676	0.0272	0.0118	0.0558	0.0553	0.0465	0.0783	$\cdots$

Figure 4.6: GMRF model as *prior* pdf for complex-valued images with analyzing window size of 16×16 pixels and model order=5. The value of texture parameters in the table are the $\mid \theta_R + j\theta_I \mid$ values for center pixel in each sub-image. The visual texture separability by the θ parameters using GMRF estimated following a linear model approach for complex-valued SAR images are similar to the θ parameters captured by the GMRF applied to amplitude images computed using IEM method. However computationally this method is very fast. It takes less than a minute to process one sub-image at the presented settings, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM.

4.5 Density parameters of generalized Gaussian distribution for wavelet coefficients of SAR images

In the previous section, we considered the cases of parametric approaches, where the texture model parameters are computed for information extraction as feature descriptor. The texture model parameters works in the Bayesian framework with assumption of a texture model (e.g. Gibbs random fields) as *prior* pdf and (generally) Gamma distribution as *likelihood* pdf. These methods are highly complex and therefore are very computationally expensive.

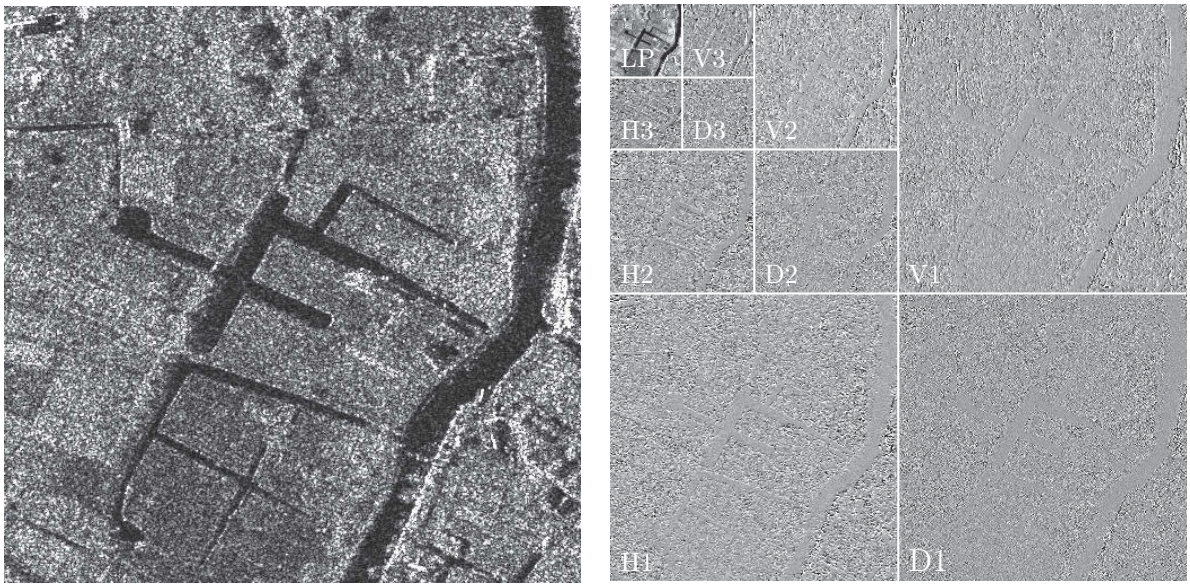
In the quest for a simpler and faster methods, it is often proposed to use directly the pdf parameters for information extractions. But generally, pdf modeling is also sometimes very complex and the number of parameters are too small to completely represent the image content. Another approach is to transform the images into a suitable space for simpler statistical modeling. This transformation will make the statistical modeling easier in order to describe the image data with simpler models. One of such effective transformation strategy is wavelet decomposition, which has been used extensively in major compression standards (JPEG-Committee [2013]). Another advantage to use wavelets is that via wavelet decomposition, temporal resolution is captured in both the time and frequency domain. Thus, the decomposition will reveal some important information in the sub-bands of the transformation space, which is otherwise hidden in spatial-image domain. Wavelet has been used for texture synthesis as it provides multi-scale and oriented sub-bands, which essentially captures the texture information. Traditionally, energy of the wavelet sub-bands are used as the feature descriptor, which is a nonparametric strategy. But we are more concerned here to search for a parametric approach for texture synthesis.

The simplest approach to exploit this property would be to simply characterize the texture in the image via marginal distribution of their wavelet sub-band coefficients as proposed in Do & Vetterli [2002]. This will be more precise compared to using the wavelet sub-band energies. In practice, the 2-D discrete wavelet transform is computed by applying a separable filter-bank to the image (Mallat [1989]). The image is decomposed into wavelet sub-bands, using suitable wavelet decomposition such as Daubechies maximally flat orthogonal filters of length 8 (D_4 filters) (Daubechies [1992]). Using this separable filter-bank, we will obtain a multi-scale pyramidal decomposition. This conventional pyramid wavelet decomposition can be applied up-to a suitable level to obtain the oriented multi-scale sub-bands. An example for the wavelet decomposition for 3 levels is shown in the figure 4.7. A SAR image of size 512×512 pixels is used. So the lowest band will have the image size of 64×64 pixels.

In order to generate a parametric feature descriptor for the image using wavelet coefficients, we need to carry out the statistical modeling for the wavelet coefficients. The wavelet transform has a property that the SAR image, which has the Gamma distribution, has generalized Gaussian (GG) distribution in the wavelet domain (Daubechies [1992]). It has also been demonstrated in Van de Wouwer *et al.* [1999] by fitting the estimated pdf curve with the actual histogram of the coefficients to model the wavelet coefficients of texture images with GG distribution. The pdf for GG distribution for a random variable $x \in (-\infty, +\infty)$ is written as follows:

$$p_x(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \mathrm{e}^{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}}, \quad \alpha, \beta \in \mathbb{R}^+, x \in \mathbb{R}. \tag{4.58}$$

With, $\Gamma(\cdot)$ as the Gamma function; μ is the expected value of the distribution; α, and β are the distribution parameters, also denoted as *scale* and *shape* parameters, respectively. The values of the *shape* parameter $\beta = 1$ and $\beta = 2$ lead to special cases of Laplacian and Gaussian distributions, respectively. Except for the lowest band or the low-pass coefficients, which corresponds to the approximation of the image, the GG distribution parameters α and β are computed for



SAR image | Wavelet decomposition to 3 levels

Figure 4.7: Various oriented sub-bands in wavelet decomposition at 3 frequency scales, on which later parameter estimation is employed. H: Horizontal frequency scale, V: Vertical frequency scale, D: Diagonal frequency scale.

all other sub-bands. In this way the statistical and multi-scale view on texture is combined. For 3 levels, we will obtain a feature descriptor of length $2 \times 9 = 18$.

4.5.0.2 Density parameter estimation using maximum likelihood approach

The parameter estimation for the GG distribution for wavelet coefficients can be carried out using the maximum likelihood (ML) approach presented in Do & Vetterli [2002], or using a globally convergent and consistent method for parameter estimation proposed in Song [2006]. We will just briefly outline the ML approach here, which we have also used in our experiments. For the ML estimation of the parameters, we define a log-likelihood function for the GG distribution with two parameters α and β as follows:

$$L(x;\alpha,\beta) = \log \prod_{i=1}^{L} p(x_i;\alpha,\beta). \tag{4.59}$$

Some more theoretical background over the ML estimation of density parameters has been presented in Appendix-C. For GG distribution this maximum likelihood estimator will have the following form (Varanasi & Aazhang [1989]):

$$\frac{\partial L(x;\alpha,\beta)}{\partial \alpha} = -\frac{L}{\alpha} + \sum_{i=1}^{L} \frac{\beta \mid x_i \mid^{\beta} \alpha^{-\beta}}{\alpha} = 0, \tag{4.60}$$

$$\frac{\partial L(x;\alpha,\beta)}{\partial \beta} = \frac{L}{\beta} + \frac{L\Psi(1/\beta)}{\beta^2} - \sum_{i=1}^{L} \left(\frac{\mid x_i \mid}{\alpha}\right)^{\beta} \log\left(\frac{\mid x_i \mid}{\alpha}\right) = 0. \tag{4.61}$$

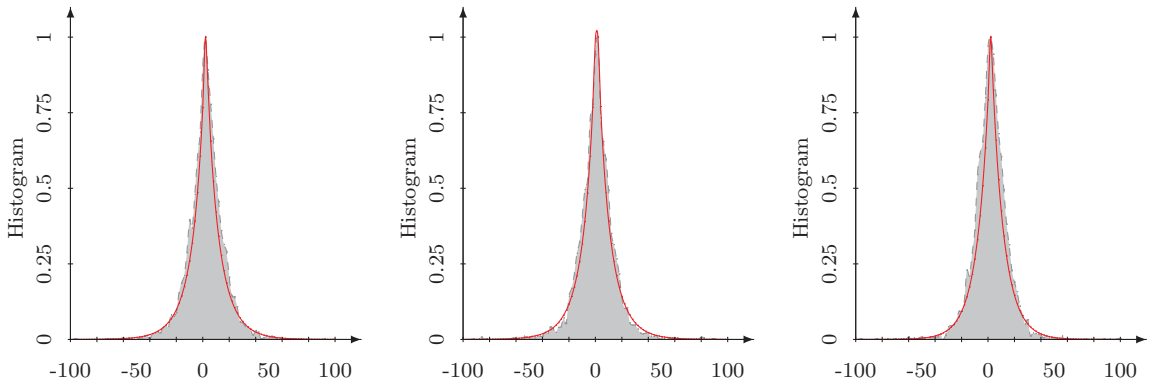
With, $\Psi(\cdot)$ as the digamma function. With *shape* parameter $\beta > 0$, the equation 4.60 will have a unique, real and positive solution for the following *scale* parameter estimate $\hat{\alpha}$:

$$\hat{\alpha} = \left(\frac{\beta}{L} \sum_{i=1}^{L} | x_i |^{\beta} \right)^{1/\beta}. \tag{4.62}$$

Substitution of the equation 4.60 into the equation 4.61, the estimate for the *shape* parameter $\hat{\beta}$ can be obtained by finding the solution of following the equation (Do & Vetterli [2002]):

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{i=1}^{L} | x_i |^{\hat{\beta}} \log | x_i |}{\sum_{i=1}^{L} | x_i |^{\hat{\beta}}} + \frac{\log \left(\frac{\hat{\beta}}{L} \sum_{i=1}^{L} | x_i |^{\hat{\beta}} \right)}{\hat{\beta}} = 0. \tag{4.63}$$

Solution for the above equation can be obtained numerically, such as using Newton-Raphson iterative procedure. We demonstrate the fitting of the GG distribution to the wavelet coefficients in the figure 4.8, with parameters estimated from ML approach.



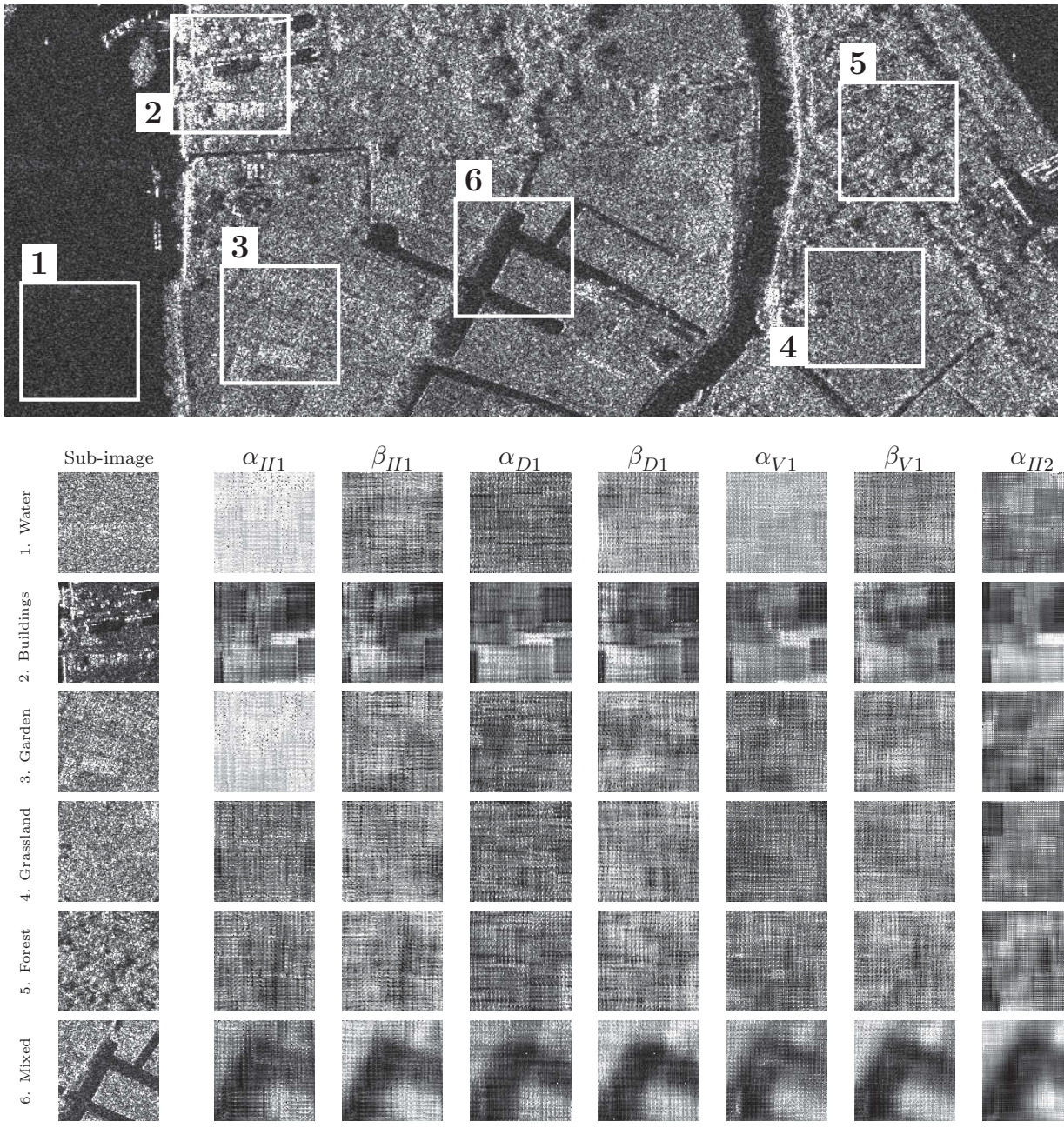
(a) H3, $\alpha = 100.9418, \beta = 0.8826$ (b) V3, $\alpha = 100.8142, \beta = 0.8706$ (c) D3, $\alpha = 96.0567, \beta = 0.9662$

Figure 4.8: Normalized histogram of the highest frequency scale sub-bands (H3, V3, and D3) fitted with the generalized Gaussian distribution. Scale and shape parameters (α, and β) for each sub-band are also presented.

Only the highest frequency scale sub-bands (H3, V3, and D3) in a wavelet decomposition to 3 levels for the image shown in the figure 4.7 are presented. The shaded gray area is the histogram, and red-line is the experimental pdf for the GG distribution along with the corresponding α and β estimates in the figure 4.8. The corresponding density parameter estimates for all the sub-bands (except the lowest sub-band which is an approximation of the image) are computed. For number of levels = 3, the parametric feature descriptor will have following form with length 18:

$$\boldsymbol{\theta}_{\text{W-GGD}} = \{\alpha_{H1}, \beta_{H1}, \alpha_{D1}, \beta_{D1}, \alpha_{V1}, \beta_{V1}, \cdots, \alpha_{V3}, \beta_{V3}\}. \tag{4.64}$$

Here, $\boldsymbol{\theta}_{\text{W-GGD}}$ stands for generalized Gaussian distribution (GGD) model parameters for wavelet coefficients.



Sub-image	β_{H1}	α_{H1}	β_{D1}	α_{D1}	β_{V1}	α_{V1}	β_{H2}	α_{H2}	β_{D2}	α_{D2}	$\cdots$
1.	1.3632	31.345	1.3854	24.624	1.3012	19.024	2.1829	29.566	1.6986	20.962	$\cdots$
2.	0.6418	64.488	0.5170	37.032	0.6606	37.868	0.6638	31.150	0.7195	37.475	$\cdots$
3.	1.2757	104.40	1.0985	101.35	1.0070	66.759	1.8911	106.74	1.7301	94.119	$\cdots$
4.	1.1336	71.996	1.2930	82.682	1.7064	77.607	1.7644	70.647	2.9260	97.649	$\cdots$
5.	1.4782	127.88	0.6697	45.164	1.0148	70.327	1.3466	90.856	1.2971	74.625	$\cdots$
6.	0.8182	23.151	0.6191	13.933	0.6043	10.736	0.8604	20.210	0.7546	14.324	$\cdots$

Figure 4.9: Generalized Gaussian distribution parameters for wavelet coefficients. Wavelet decomposition is carried out up-to 2 levels and analyzing window size used is 32×32 pixels. The density parameter estimation in wavelet space using maximum likelihood approach are relatively fast. It takes around 1 minutes to process one sub-image at presented settings, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM. However, the texture separability is relatively poor.

4.6 Summary

The following points have been discussed in this chapter.

- The parametric information extraction approaches are methods where a specific model for the given data set is assumed. The model parameters are used as a descriptor to describe the image content. The parametric information extraction approaches for spatial content understanding can be further classified into approaches applied to images in spatial domain and approaches applied to images within a transformation domain.

- The parametric approaches applied to images in spatial domain generally refers to the application of Bayesian inference to SAR images which is a consequence of the multiplicative speckle model. An introduction to two levels of Bayesian inference has been presented. The knowledge of the *prior* pdf, which is assumed to be the model of texture of the underlying backscatter reflectance, helps in improving the accuracy of the parameter estimation.

- For parametric approaches applied in spatial domain, generally Gibbs based *prior* models, such as Gauss-Markov random fields (GMRF) and autobinomial (ABM) model are used for texture synthesis in amplitude SAR images. Theoretical understanding of these models has been presented. Guidelines for maximum *a posteriori* (MAP) parameter estimation following an iterative evidence maximization approach has been presented along with preliminary results of estimated parameters which will be later used for performance evaluations.

- We also extend the GMRF *prior* pdf to complex-valued SAR images. For the parameter estimation of complex-valued SAR images, we proposed to follow a linear model approach by taking the advantage of the symmetry and linearity of the model. With Gauss-Markov random fields applying to the complex-valued SAR images, we dwell into the statistical texture analysis of the complex-valued SAR images, which is largely limited to the amplitude images until now.

- For parametric approaches applied to images within a transformation domain, we consider the density parameters of generalized Gaussian distribution for wavelet coefficients of SAR images. The inclusion of this method in our analysis is attempted to provide the evidence of the advantage of transformation of image pixels, at least for the patch-oriented image categorization application which will be evident in the analysis presented in Part-III.

Chapter 5

Information extraction in SAR images: nonparametric approaches

Parametric approaches deal with the problem of estimating a set a parameters, by assuming a particular model for the set of given pixels in an image. Contrary to this, nonparametric approaches are not postulating any particular model for a set of pixels in image. The objective of this chapter is to discuss some of the selected and a newly proposed nonparametric method for information extraction, with introductory remarks about the nonparametric approaches in the first section.

The nonparametric analysis of images can be carried out in 'spatial domain' by viewing images as a collection of pixels, or in 'spatial-frequency domain' by viewing images as the sum of sinusoids of infinite extent. The analysis using the gray level co-occurrence matrices (GLCM) is a method in 'spatial domain'. The texture parameter estimation in GLCM is based on the computation of a set of spatial-dependence probability distribution matrices, which is discussed in the second section of the chapter.

The other approach in nonparametric analysis is frequency analysis method which can be carried out in the 'spatial-frequency domain'. This leads us to introduction to a non-linear spectral feature descriptor (NL-SFD) computed on the spectrum of the SAR images, which is discussed in the third section.

However, it has been observed that both the cases (GLCM and NL-SFD) are two opposite extremes of a continuum of possible joint space/spatial-frequency representation. In real life signals, such as SAR images, it is useful to carry out the analysis of signals in joint space/spatial-frequency representation, which is also popularly known as the joint time-frequency analysis (JTFA), such as using the fractional Fourier transform (FrFT). The FrFT is a generalization of the Fourier representation. The FrFT is a type of JTFA where a complex-valued signal is decomposed in terms of chirps. Another type of JTFA is the Gabor expansion, *which is based on a filter-band paradigm. The FrFT-based method and Gabor-based method are discussed in third and fourth section of this chapter, respectively. We will close this chapter with a short summary in the last section.*

5.1 Introduction to nonparametric approaches

An approach for information extraction and image processing applications which is widely applicable in multimedia images as well as remote-sensing images is the nonparametric analysis. Actually the definition of nonparametric analysis is not very well defined. In contrast to the parametric approaches which provide few parameters for the texture / pattern characterization by postulating a certain model for the data, the nonparametric approaches may be classified into a group of methods which provide a large number of *kind-of-parameters* without postulating a certain model for the data. This provides more flexibility as the internal architecture parameters are to be set by the user. The nonparametric approaches can be classified into two classes, first is the histogram-based method and second is based on the statistical measures of intermediate sub-images from the original image.

5.1.1 Histogram-based methods

The first class of nonparametric approaches uses the low-level feature, i.e. the histogram (or color histogram) as a stable representation of the objects and texture in an image. The histogram of the image serves as the feature descriptor which is very large in size. In case of gray level images, naturally the size of feature descriptor (similar to the histogram) is equal to the number of gray levels in the image. In case of colored images, a color histogram is used in similar fashion. This large feature descriptor is given as input to the standard classification methods (unsupervised such as k-Means, supervised such as k-nearest neighbor), or other sophisticated classification routines such as artificial neural networks (ANN) or support vector machines (SVM). Indexing colored multimedia images using the color-histogram has been presented in Swain & Ballard [1990] and the use of histogram based segmentation in SAR images has been demonstrated in Lee & Jurkevich [1989]. For the estimation of probability density function (pdf), several nonparametric kernel-based estimation and regression architectures have been proposed in the literature that have proved to be effective estimation tools, such as standard Parzen window estimators (Parzen [1962] Moser *et al.* [2006]).

5.1.2 Sub-images based methods

The second class of nonparametric approaches generates some intermediate sub-images and statistical measure on these sub-images are used to describe the image content. The size of feature descriptor (which consists of the statistical measure of the sub-images) in this class is much smaller than the histogram-based nonparametric approaches. The general architecture of the nonparametric analysis is presented by a block diagram shown in the figure 5.1. To obtain a feature descriptor for an image, popularly a two stage process is employed. The first stage refers to the generation of the sub-images. In the second stage, feature extraction process is applied to the obtained sub-images. Some of the steps in various stages in the figure 5.1 may be ignored in some methods. In this section, we will illustrate only the general methodological framework, keeping the SAR satellite images as our interest domain.

In the first stage, sub-images are generated from the original SAR images. The images obtained from the SAR satellites are inherently the complex-valued images, but generally (not always) feature extraction techniques only employ the use of amplitude SAR images. In the first step, the input image $I_{(M\times N)}$, (amplitude or complex-valued, as the case may be) is first preprocessed to obtain an approximate image $AI_{(M_h\times N_h)}$. The preprocessing steps may involve reduction in the geometrical resolution (down-sampling) or reduction in the radiometrical resolution (quantization) etc. Representing the size of original image as $M \times N$ pixels, with M

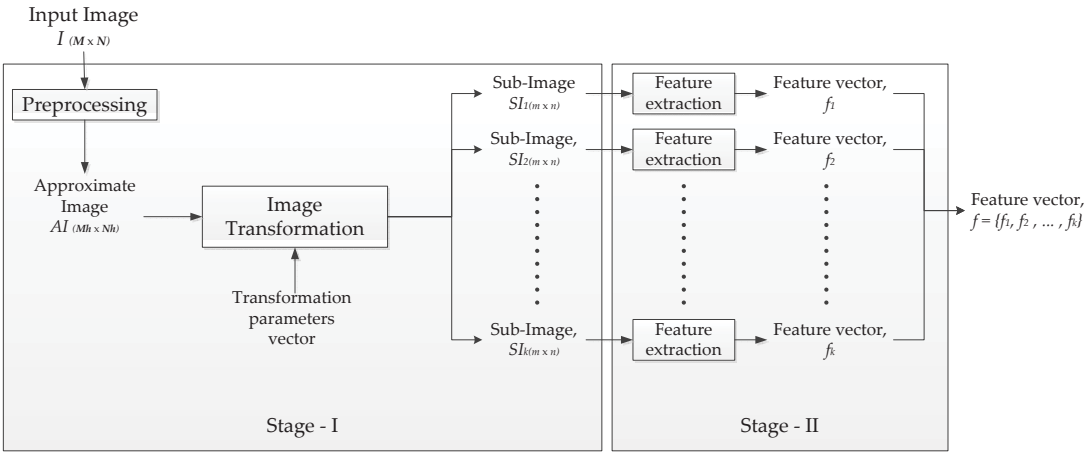


Figure 5.1: Architecture of the sub-images based nonparametric analysis of images for information extraction.

as the number of rows and N as the number of columns in the image, the approximate image has the size of $M_h \times N_h$ pixels, again with M_h as the number of rows and N_h as the number of columns. The size $M_h \times N_h$ pixels is typically smaller than or equal to the size $M \times N$ pixels. This preprocessing step is employed to reduce the computational burden in next step without compromising the overall accuracy of performance. The next step is the image transformation. In this step, the image pixels are transformed into a suitable transformation domain, thus representing them in various planes of this transformation domain in form of sub-images $SI_{k(m\times n)}$, with $k = 1, 2, \cdots, k$. The advantage of this transformation is to reveal the texture / pattern specific information in different planes. The transformation parameter vector decides the number k of generated sub-images. The size $m \times n$ pixels of sub-images is generally smaller than or equal to the input image, i.e. the approximate image $AI_{(M_h\times N_h)}$. These sub-images may be real-valued or complex-valued. The transformation process is usually the representation of the spatial-domain image into frequency domain, or joint time-frequency domain in various sub-bands. The other approach could be by applying various filter-banks (e.g. wavelet/ Gabor filters) to obtain filtered sub-bands. Another approach is to generate base-images (or base-matrices) with taking into consideration the interaction between neighboring pixels in certain orientation, e.g. in gray level co-occurrence matrix (GLCM) based methods.

Once the sub-images $SI_{k(m\times n)}$ have been generated in the first stage, a feature extraction methodology is applied to the sub-images in the second stage. Normally, the features extracted from these sub-images are the statistical sample moments up to certain order. However, in Julesz [1981], it has been demonstrated that only linear mechanisms are not sufficient for texture discrimination. So when the sub-images are originally complex-valued, a non-linearity is introduced by computing the amplitude sub-images before texture feature extraction. In other cases, when the sub-images are real-valued (e.g. in the case of GLCM based methods), non-linear feature for these sub-images are computed. The feature descriptors f_k, with $k = 1, 2, \cdots, k$, obtained from sub-images, are sequentially concatenated to formulate the final feature descriptor $f = \{f_1, f_2, \cdots, f_k\}$ for the input image $I_{(M\times N)}$. When the sub-images are highly correlated with each other, a feature selection scheme may be employed: Alternatively an isotropic feature descriptor may be generated using the individual feature descriptors from sub-images. In this thesis we restrict our analysis to the the feature descriptor formulation shown in the figure 5.1.

The sub-images can be generated directly in the image domain by considering the interaction between the pixels in its neighborhood. One example of such methods is the GLCM, which will be analyzed in this thesis as a nonparametric approach for image domain. The other approach is to transform the images into a suitable space. The projection of the image on various planes in multi-dimensional transform space will be various sub-images. The examples of such transformation can be Fourier transform, the fractional Fourier transform (FrFT), and Gabor filtering (a kind of wavelet decomposition) etc. These methods will be analyzed in this thesis as nonparametric approaches for image within a transformation space.

Unlike parametric approaches where the model parameters (texture or density parameter) is used as feature descriptor, different methodologies are needed to be adapted to generate a feature descriptor in nonparametric approaches. In the following we present the theoretical background on the considered methods and guidelines to form a feature descriptor to describe the spatial content of the images.

5.2 Gray-level co-occurrence matrix

The GLCM or the gray-tone spatial-dependence matrix based method is a way to extract second-order statistical texture features used in various image processing applications proposed by Haralick in the 1970*s* (Haralick *et al.* [1973], Haralick [1979]).

In Baraldi & Parmiggiani [1995], the use of GLCM's texture parameters for the texture characterization in optical satellite images (NOAA AVHRR) have been investigated. It has also been commented that in certain cases gray level difference histogram (GLDH) outperforms the GLCM when considering the overall texture measurement accuracy, along with computer storage and computation time. The texture features generated using the GLCM have been successfully applied to the texture analysis for mapping sea ice patterns in the low-resolution ERS-1 SAR images in Soh & Tsatsoulis [1999]. The combined use of the GLCM texture parameters with Markov random fields (MRFs) texture features have been suggested to improve the classification of SAR sea-ice imagery in Clausi [2000], later extensive comparative study of GLCM and MRFs based texture features have been presented in Clausi & Yue [2004]. The use of GLCM in efficient texture analysis in ENVISAT-ASAR images by introducing the notion of patch re-occurrences has been proposed and compared with the texture features obtained using Gabor expansion in Kandaswamy *et al.* [2005]. The texture features based on GLCM generally outperforms other classical methods of classification, at least in the case of low-resolution SAR and optical satellite images. For SAR images, the use of GLCM is largely limited to the sea-ice mapping and clustering application only. We will investigate the usability of GLCM to the land-cover classification as well as to the patch-oriented image categorization to the very-high-resolution SAR images only.

5.2.1 Computation of gray level co-occurrence matrix

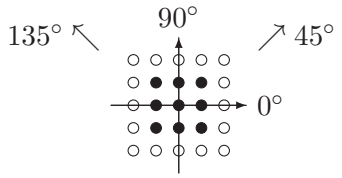
The computation of the texture parameters from the GLCMs have been extensively discussed in literature. Without going into much details in this section, we will briefly recall the theoretical overview of the generation of the GLCMs and extraction of the textural parameters for SAR image

The GLCMs $P_{i,j|d,\theta}$ have the number of rows and columns equal to the number of gray levels, G, in the image. To generate texture features based on the co-occurrence probabilities, statistics are applied to the probabilities (Clausi [2000]). The $(i,j)^{th}$ entry in the matrices of gray-tone spatial-dependence frequencies, $P_{i,j|d,\theta}$ at orientation θ°, and offset d, is the statistical

probability values of changes between gray tone i and gray tone j at a particular orientation angle θ and a particular offset d. An example of computation of the GLCMs for an analyzing window size of 4×4 pixels is demonstrated in the figure 5.2. Here the range of gray levels is $0-3$, i.e. the size G of the GLCMs $P_{i,j|d,\theta}$ (figure 5.2c) will be 4×4 pixels. Each element (i, j) of the GLCMs contains the number of total gray levels with respective i and j neighbors in the analyzing window for a particular orientation. In this example, we consider the offset value $= 1$ for four orientation, i.e. 0°, 45°, 90°, and 135°, as shown in the figure 5.2b. This computation in each orientation results into 4 GLCMs, as shown in the figure 5.2d.

0	1	1	3
1	0	3	0
0	0	1	2
1	2	3	0

(a) Example of an 4×4 analyzing window, with gray level quantized to 4 levels.



(b) Geometry of measurement of gray level co-occurrence matrix for offset $= 1$ and four orientations.

Gray values	0	1	2	3
0	$\#_{(0,0)}$	$\#_{(0,1)}$	$\#_{(0,2)}$	$\#_{(0,3)}$
1	$\#_{(1,0)}$	$\#_{(1,1)}$	$\#_{(1,2)}$	$\#_{(1,3)}$
2	$\#_{(2,0)}$	$\#_{(2,1)}$	$\#_{(2,2)}$	$\#_{(2,3)}$
3	$\#_{(3,0)}$	$\#_{(3,1)}$	$\#_{(3,2)}$	$\#_{(3,3)}$

(c) A general form of the gray level co-occurrence matrix for a particular orientation, taking the quantization level to 4 gray levels. Here, $\#_{(i,j)}$ denotes the number of gray levels with respective i and j neighbors in the analyzing window for a particular orientation.

$$P_{0^\circ} = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 0 \end{pmatrix}, P_{45^\circ} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, P_{90^\circ} = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{pmatrix}, P_{135^\circ} = \begin{pmatrix} 1 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

(d) Computation of gray level co-occurrence matrices for four orientation: P_{0°,P_{45°,P_{90°,and P_{135°,for the analyzing window shown in (a), Later, the texture parameters have to be estimated for these matrices.

Figure 5.2: Example of generation of gray level co-occurrence matrices. The size of analyzing window is 4×4, the number of gray level after quantization is 4 , and four orientations, i.e. 0°, 45°, 90°, and 135° have been used

In practice, the number of gray levels is significantly reduced by quantization in the computation of the GLCMs. In case we are supposed to use the full gray level range in an image, the matrices on which texture parameters needs to be computed will be so voluminous that the usability of this method for practical applications will not be feasible any more due to high computational costs. For example, the radiometric resolution of TerraSAR-X images being 16 bits, will provide a GLCM of size $2^{16} \times 2^{16}$, with $G = 2^{16}$. Whereas, even visually, quantization into 2^4 gray levels is sufficient to discriminate the textures (Albgretsen [2008]). Thus in our study concerning with the very-high-resolution TerraSAR-X images, we have also quantized the images to $G = 16$ gray levels. There has been no rigorous study available which provide the optimal selection of orientation angle θ, but in literature the use of four orientations, i.e. 0°,

45°, 90°, and 135° has been advocated with demonstration of sufficient accuracy of classification at these orientation angles (Barber & Ledrew [1991], Clausi & Yue [2004]). The offset value d used is 4 pixels, which has been selected empirically. These settings will provide us a GLCM P_θ of size 16×16 for each orientation θ. On the resulting GLCMs, P, 12 parameters are computed as summarized in Albgretsen [2008], which are presented in next part of this section.

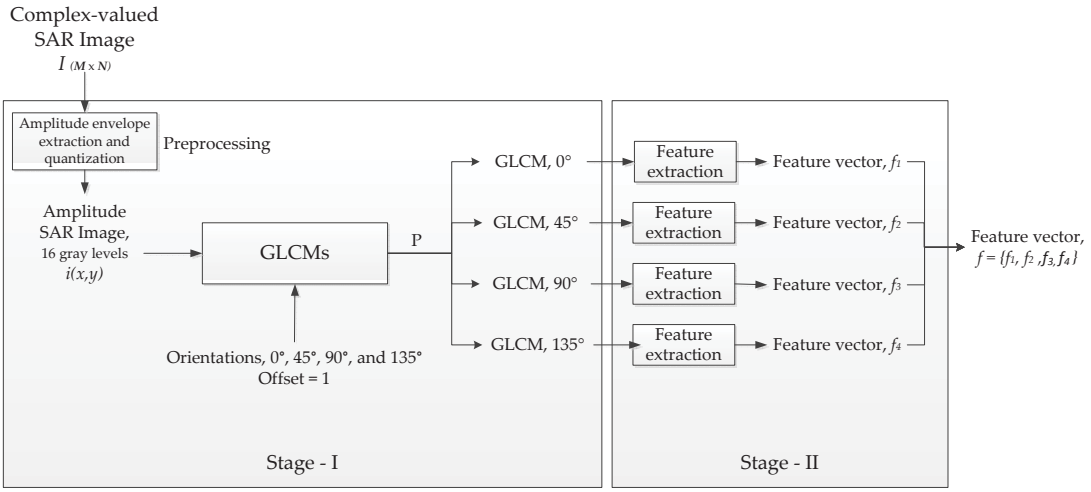


Figure 5.3: Architecture of the sub-images based GLCM nonparametric analysis of images for information extraction.

5.2.2 Feature descriptor based on the gray level co-occurrence matrix

On the GLCMs $P_{i,j|d,\theta}$, with $G = 16$, following texture features have been computed. The first and second statistical moments (mean and variance) are computed on gray level co-occurrence matrices $P_{i,j|d,\theta}$ as follows:

$$1.\ \text{Mean:}\quad \mu = \frac{1}{\mathrm{G} \times \mathrm{G}} \sum_{i=1}^{\mathrm{G}} \sum_{j=1}^{\mathrm{G}} \mathrm{P}_{i,j}, \qquad 2.\ \text{Variance:}\quad \sigma^2 = \frac{1}{\mathrm{G} \times \mathrm{G}} \sum_{i=1}^{\mathrm{G}} \sum_{j=1}^{\mathrm{G}} (\mathrm{P}_{i,j} - \mu)^2$$

In Albgretsen [2008], it was suggested to compute mean and variance in x and y direction separately, whereas our experiments advocate that combined computation of mean and variance in both direction will reduce the size of feature descriptor without compromising the classification accuracy.

Next element in the feature descriptor is the entropy, which is a standard measure in texture computation. Entropy suggests the degree of randomness and uncertainty associated with the pixels in the image. As the elements of the GLCMs $P_{i,j|d,\theta}$ are actually the statistical probability values of the gray level change in the image, thus entropy can be computed as follows:

$$3.\ \text{Entropy:}\quad -\sum_{i=1}^{\mathrm{G}} \sum_{i=1}^{\mathrm{G}} \mathrm{P}_{i,j} \log \mathrm{P}_{i,j},$$

The angular second moment (ASM) or energy as well as the inverse difference moment (IDM) are the measures of homogeneity of the image. The value of ASM will be high in case of a homogeneous scene due to the relatively high contribution from some of the gray levels. The

weighting factor $1+(i-j)^2$ in IDM will result a low IDM value of inhomogeneous areas. The ASM and IDM are computed as follows:

$$4.\ \text{ASM:}\quad \sum_{i=1}^{\mathrm{G}}\sum_{i=1}^{\mathrm{G}}\mathrm{P}_{i,j}^2, \qquad 5.\ \text{IDM:}\quad \sum_{i=1}^{\mathrm{G}}\sum_{i=1}^{\mathrm{G}}\frac{\mathrm{P}_{i,j}}{1+(i-j)^2}. \tag{5.1}$$

Correlation between the elements of the GLCM will provide the relative linear dependencies of the pixels in image at specific positions, which is computed as follows:

$$6.\ \text{Correlation}: \quad \sum_{i=1}^{\mathrm{G}}\sum_{j=1}^{\mathrm{G}}\frac{ij\mathrm{P}_{i,j}-\mu_i\mu_j}{\sigma_i\sigma_j}.$$

Here, μ_x, μ_y, σ_x and σ_y are the first and second statistical moments of the marginal-probability matrices P_x and P_y. With, $P_x(i)=\sum_{j=1}^{G}\mathrm{P}(i,j)$ and $\mathrm{P}_y(j)=\sum_{i=1}^{G}\mathrm{P}(i,j)$. Contrast and dissimilarity measures the local intensity variations which favors the contributions from non-diagonal elements of GLCM, computed as follows:

$$7.\ \text{Contrast:}\quad \sum_{i=1}^{\mathrm{G}}\sum_{i=1}^{\mathrm{G}}\mathrm{P}_{i,j}(i-j)^2, \qquad 8.\ \text{Dissimilarity:}\quad \sum_{i=1}^{\mathrm{G}}\sum_{i=1}^{\mathrm{G}}\mathrm{P}_{i,j}|i-j|^2.$$

The last two features, cluster shade (CS) and cluster prominence (CP) are measures of the skewness of the GLCM, i.e. they are the measure of lack of symmetry (Ion [2009]), and computed as follows:

$$9.\ \text{CS:}\quad \sum_{i=1}^{\mathrm{G}}\sum_{i=1}^{\mathrm{G}}(i+j-\mu_i-\mu_j)^3\mathrm{P}_{i,j}, \qquad 10.\ \text{CP:}\quad \sum_{i=1}^{\mathrm{G}}\sum_{i=1}^{\mathrm{G}}(i+j-\mu_i-\mu_j)^4\mathrm{P}_{i,j}$$

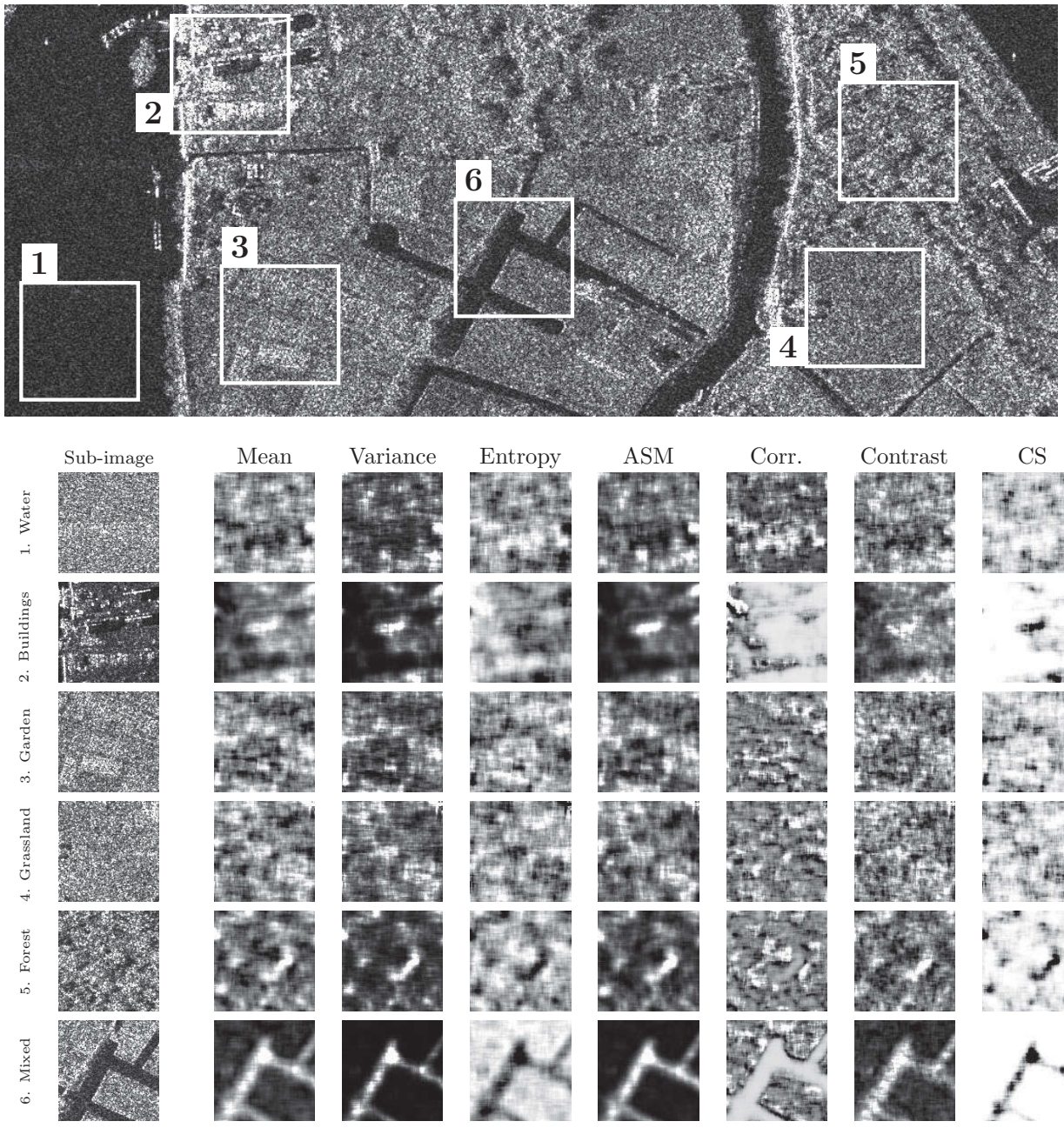
It can be seen that many of the features are strongly correlated with each other, which need to be computed for each orientation separately. Thus a feature selection criterion can be adapted before classification procedure. Haralick proposed the use of angular mean, $M_{f_{\mathrm{GLCM}}}(d)$ and range $R_{f_{\mathrm{GLCM}}}(d)$ of all the presented features f_{GLCM} at d offset and N_θ orientations for reducing the size of feature descriptor to input to a classifier. Angular mean and range can be computed as follows (Haralick *et al.* [1973], Albgretsen [2008]):

$$M_{f_{\mathrm{GLCM}}}(d)=\frac{1}{N_\theta}\sum_{\theta}f_{\mathrm{GLCM}}(d,\theta), \tag{5.2}$$

$$R_{f_{\mathrm{GLCM}}}(d)=\max_{\theta}[f_{\mathrm{GLCM}}(d,\theta)]-\min_{\theta}[f_{\mathrm{GLCM}}(d,\theta)]. \tag{5.3}$$

However, our analysis suggests to use all the features in one feature descriptor for better classification results. Using four options of orientation, 0°, 45°, 90° and, 135°, provides us a feature descriptor of length 40 (as shown with an example in the figure 5.4):

$$\bar{f}_{\mathrm{GLCM}}=\{f_{\mathrm{GLCM}_{0^\circ}\{1\text{ to }10\}},f_{\mathrm{GLCM}_{45^\circ}\{1\text{ to }10\}},f_{\mathrm{GLCM}_{90^\circ}\{1\text{ to }10\}},f_{\mathrm{GLCM}_{135^\circ}\{1\text{ to }10\}}\}. \tag{5.4}$$



Sub-image	Mean	Variance	ASM	Corr.	Contrast	Dissi.	CS	CP	···
1.	6.5077	19.165	6.18e-3	1.86e-2	3.67e+1	4.6709	2.09e+2	4.74e+3	···
2.	27.179	1923.9	3.45e-1	-1.49e-4	1.76e+2	21.793	-3.14e+5	1.36e+7	···
3.	9.7577	30.096	1.06e-2	-2.66e-2	5.14e+1	6.4924	-8.11e+2	1.24e+4	···
4.	10.714	49.768	1.53e-2	-1.14e-2	5.49e+1	7.1250	-1.77e+3	2.87e+4	···
5.	8.8163	26.039	1.06e-2	-1.65e-2	4.78e+1	5.9719	-3.42e+2	6.48e+3	···
6.	10.982	85.314	5.27e-2	-6.02e-3	8.88e+1	10.809	-3.18e+3	5.14e+4	···

Figure 5.4: Gray level co-occurrence matrix based feature descriptor. The images have been first quantized to 16 gray levels. GLCMs using four options of orientation and pixel offset 1 have been computed before feature descriptor generation. The feature separability is good in several regions, but due to quantization step overall accuracy will probably be poor. It takes around 5 minutes to process one sub-image at presented settings, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM.

5.3 Non-linear feature descriptor based on the Fourier spectrum

The analysis of the images can be carried out in a transformation space, such as spatial-frequency domain using the method of Fourier representation. Any real or complex-valued continuous signal $x(t)$ defined on the interval (t_0, t_0+T) can be represented by following expansion (Ahmed & Rao [1975]):

$$x(t) = \sum_{n=-\infty}^{\infty} c_n u_n(t), \tag{5.5}$$

with, c_n as the n^{th} coefficient of the expansion and $u_n(t)$ as the set of real or complex-valued continuous basis functions. In case of Fourier representation of a periodic signal, the set of functions $\{u_n(t)\}$ is the exponential signals $\{e^{jn\omega_0 t}, n =, 0, \pm 1, \pm 2, \cdots\}$. Here ω_0 is the fundamental angular frequency, related to the time period T_p as $\omega_0 = 2\pi T_p$. Thus the Fourier series expansion for a continuous time periodic signal is given as follows:

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t}, \text{with} \quad c_n = \frac{1}{T_p}\int_{T_p} x(t)e^{-jn\omega_0 t}dt. \tag{5.6}$$

The coefficients of expansion c_n are termed as the Fourier coefficients, which are complex-valued. The above expansion is possible as long as the *Dirichlet conditions* with strengthened requirements are satisfied (Gaskill [1978]), i.e. the signal $x(t)$:

- has finite number of discontinuities in any period,
- has a finite number of maxima and minima in any period,
- is absolutely integrable on the interval $(-\infty, \infty)$.

The transition from Fourier series to Fourier transform happens when we allow period to increase without limit, i.e. signal becomes aperiodic. Thus the spectrum of $x(t)$ is obtained by taking the limit in the equation 5.6 as $T_p \to \infty$. Thus the Fourier transform pair for a continuous time aperiodic signal $x(t)$are given as:

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega, \text{and} \quad X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt. \tag{5.7}$$

Again the following *Dirichlet conditions* guarantee the existence of the Fourier transform:

- has finite number of discontinuities,
- has a finite number of maxima and minima,
- is absolutely integrable.

Up to now we discussed the case of continuous time or analog signals. In this case of continuous time periodic signals, Fourier series representation can consist of an infinite number of frequency components ranging from $-\infty$ to $+\infty$ and frequency spacing between two successive harmonically related frequencies is $1/T_p$ (John G. Proakis [2006]). However, in the real-life situations, we encounter the case with discrete time signals where the frequency range is unique over the interval $(-\pi, \pi)$ or $(0, 2\pi)$. Consequently, the Fourier series representation for a discrete time periodic signal will contains at most N frequency components, given as follows:

$$x(n) = \sum_{m=0}^{N} c_n e^{j2\pi mn/N}, \text{with} \quad c_m = \frac{1}{N}\sum_{n=0}^{N} x(n)e^{-j2\pi mn/N}. \tag{5.8}$$

Where N is the fundamental period. Similar to the case of continuous time periodic signals, the frequency analysis of discrete time periodic signals can be extended to the aperiodic case. As now the signal is discrete in time, integral is replaced with summation. Moreover, the frequency range is limited to $(-\pi, \pi)$ or $(0, 2\pi)$. Thus Fourier transform pair can be written as follows:

$$x(n) = \frac{1}{2\pi} \int_{2\pi} X(\omega) \mathrm{e}^{j\omega n} d\omega, \text{and} \quad X(\omega) = \sum_{n=-\infty}^{\infty} x(n) \mathrm{e}^{-j\omega n}. \tag{5.9}$$

Above formulations have been presented for the case of one-dimensional. These formulations can be extended for a two-dimensional case $x(r, a)$ by carrying out a sequential transform in both directions, e.g. for a SAR image, with two variables r and a defining the range and azimuth bins in spatial domain. The Fourier transform pair for a SAR image $x(r, a)$ with two variables of dimension $M \times N$ can be written as follows:

$$x(r,a) = \frac{1}{N^2} \sum_{k=0}^{M} \sum_{l=0}^{N} X(u,v) \mathrm{e}^{j2\pi\left(\frac{kr}{M} + \frac{la}{N}\right)}, \quad X(u,v) = \sum_{k=0}^{M} \sum_{l=0}^{N} x(r,a) \mathrm{e}^{-j2\pi\left(\frac{ku}{M} + \frac{lv}{N}\right)} \tag{5.10}$$

The Fourier transform of the image $x(r, a)$, i.e. the spectrum $X(u, v)$ is used for the generation of the feature descriptor.

5.3.1 A non-linear feature descriptor for SAR images based on Fourier transform

The use of spectrum in the analysis of speech signal has been extensively studied in literature. Spectrum based standard features have been studied for speech recognition in Hunt *et al.* [1980] and for music-speech discrimination in Scheirer & Slaney [1997]. Later these features have been revisited in Tzanetakis & Cook [2002], and Li & Ogihara [2006]. We tried to adapt these spectral features for the texture and structure analysis in the SAR images. We presented the use 7 features motivated from the mentioned *timbre texture features* in Singh *et al.* [2009] for pixel-based classification in very-high-resolution SAR images. In Popescu *et al.* [2010], we demonstrated the same feature descriptor on the application to patch-oriented image categorization in very-high-resolution SAR images. An extensive study on patch-oriented image categorization has been later presented in Popescu *et al.* [2012].

No we will provide the theoretical details of these 7 non-linear features based on the spectrum of SAR images. First two features (mean and variance) are based on statistical properties and next four are motivated from *timbre texture features* used for music genre classification. The mean, μ, is the average value of the amplitude of the spectral decomposition, and the variance, σ^2, provides the information about the statistical dispersion of the data given as follows:

$$\mu_{\text{spectrum}} = \frac{1}{\mathrm{M} \times \mathrm{N}} \sum_{u=1}^{\mathrm{M}} \sum_{v=1}^{\mathrm{N}} |\hat{X}(u,v)|, \quad \sigma^2_{\text{spectrum}} = \frac{1}{\mathrm{M} \times \mathrm{N}} \sum_{u=1}^{\mathrm{M}} \sum_{v=1}^{\mathrm{N}} \left(|\hat{X}(u,v)| - \mu\right)^2. \tag{5.11}$$

Where, $\mathrm{M} \times \mathrm{N}$ is the size of the sub-image, which is also the size of the estimated spectrum $\hat{X}(u, v)$. Here, (r, a) are the spatial range and azimuth bins, and (u, v) are the range and azimuth frequency bins respectively. Variance is the square of σ, denoted as standard deviation.

Spectral centroid is the center of mass or centroid of the amplitude of the spectral decomposition for a signal. Spectral centroid for an image is computed in range (SC^r) and azimuth

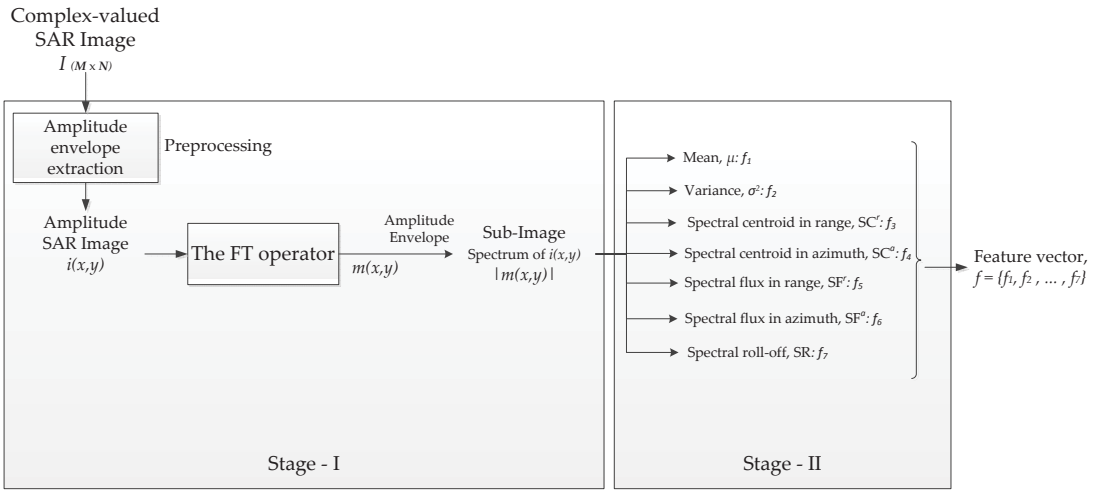


Figure 5.5: Architecture of the sub-images based spectral nonparametric analysis of images for information extraction. FT stands for the canonical Fourier transform operator.

(SC^a) directions separately as follows:

$$3.\quad \mathrm{SC}^r = \frac{\sum_{r=1}^{\mathrm{M}}\sum_{a=1}^{\mathrm{N}} |\hat{X}(u,v)| * f^r}{\sum_{r=1}^{\mathrm{M}}\sum_{a=1}^{\mathrm{N}} |\hat{X}(u,v)|}, \quad 4.\quad \mathrm{SC}^a = \frac{\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}} |\hat{X}(u,v)| * f^a}{\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}} |\hat{X}(u,v)|}. \tag{5.12}$$

Where, f^r and f^a denotes the magnitude of the frequency bin number in range and azimuth direction respectively. Spectral centroid is the measure of the brightness of the spectrum.

Spectral flux measures how the local spectrum is changing with successive steps. It is nothing but the square of the difference between magnitude of two successive spectra. Similar to spectral centroid, spectral flux is also computed in range (SF^r) and azimuth (SF^a) directions separately as follows:

$$5.\quad \mathrm{SF}^r = \sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}} (|\hat{X}(u,v)| - |\hat{X}(u-1,v)|)^2, \quad 6.\quad \mathrm{SF}^a = \sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}} (|\hat{X}(u,v)| - |\hat{X}(u,v-1)|)^2. \tag{5.13}$$

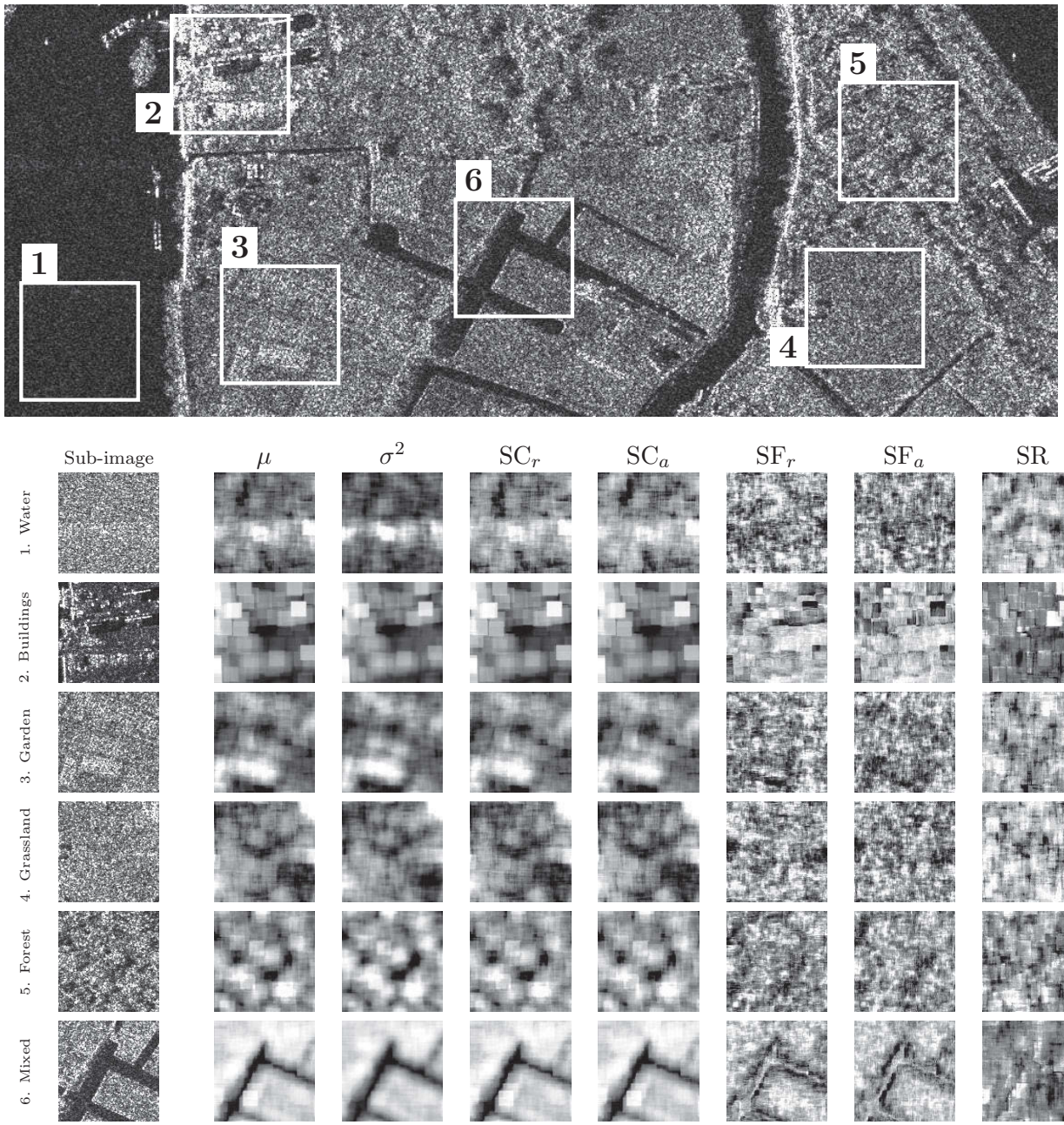
Spectral roll-off is a measure of spectrum energy and represents the frequency bins below which $\alpha\%$ of the spectrum energy is concentrated. We consider the value of $\alpha = 85$. Larger value of spectral roll-off corresponds to the higher frequencies in an image patch.

$$7.\quad \sum_{r=1}^{\mathrm{Rr}}\sum_{a=1}^{\mathrm{M}} \mathrm{SR}(r,a) = \frac{\alpha}{100}\sum_{r=1}^{\mathrm{M}}\sum_{a=1}^{\mathrm{N}} |X(u,v)|. \tag{5.14}$$

In equations 5.11 to 5.14, the subscripts of summation running from $r = 1$ to M and $a = 1$ to N are representing the range and azimuth bins respectively. We have taken M=N in all the cases in our analysis. So the contextual feature descriptor based on the full spectrum of SAR image has the following form with the length of 7 elements:

$$\bar{f}_{\text{NL-SFD}} = \{\mu, \sigma^2, \mathrm{SC}^r, \mathrm{SC}^a, \mathrm{SF}^r, \mathrm{SF}^a, \mathrm{SR}\}. \tag{5.15}$$

An example of computed non-linear spectral feature descriptor $\bar{f}_{\text{NL-SFD}}$ and visual inspection for texture separability is presented in the figure 5.6.



Sub-image	μ	σ^2	SC_r	SC_a	SF_r	SF_a	SR
1.	2.937e+2	3.117e+5	5.693e+5	5.540e+5	1.005e+0	8.227e-1	2.200e+2
2.	8.983e+2	2.575e+6	1.736e+6	1.655e+6	5.505e-1	5.497e-1	2.230e+2
3.	1.289e+3	6.748e+6	2.530e+6	2.446e+6	8.031e-1	7.655e-1	2.240e+2
4.	8.516e+2	2.889e+6	1.652e+6	1.654e+6	8.656e-1	8.933e-1	2.220e+2
5.	1.554e+3	7.358e+6	3.047e+6	2.943e+6	5.634e-1	7.378e-1	2.240e+2
6.	7.730e+2	1.442e+6	1.490e+6	1.440e+6	4.845e-1	5.615e-1	2.290e+2

Figure 5.6: Non-linear spectral feature descriptor. With visual assessment, we can expect to have a good texture / category discrimination. E.g. in sub-image 1, the roughness of sea on the lower part is distinctively visible in several features. As well as in sub-image 3, the variation in the garden texture at lower left corner is also captured. The spectral roll-off seems to behave very good in discriminating the spatial structures for man-made objects in sub-image 3. It takes less than a minute to process one sub-image at presented settings, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM.

5.4 The fractional Fourier transform

Whereas the GLCM presented in section 5.2 is a method in spatial domain, the method of Fourier representation presented in section 5.3 is dealing the signals purely in frequency domain. These are two opposite extremes of a continuum of possible joint space/spatial-frequency representation. A more detailed insight of the signal can be achieved by expanding the signals jointly in time and frequency domain, i.e. by carrying out the joint time frequency analysis (JTFA). The JTFA can be carried out using the linear transformation such as the fractional Fourier transform (FrFT), filter-bank based approaches such as Gabor expansion, or using quadratic transformation such as Wigner-Ville decomposition.

The use of Wigner-Ville decomposition for the moving-target detection and imaging in SAR images was demonstrated in Barbarossa & Farina [1992]. Being a quadratic transform, the Wigner-Ville distribution is influenced by cross-term interference when more than one target is present (Qian & Chen [1999]). In Sun *et al.* [2002], use of the FrFT for moving-target detection was proposed. The relation between the FrFT with the Wigner-Ville distribution is derived and its advantages over the Wigner-Ville distribution demonstrated. The FrFT is a linear transformation free from cross-term interference. Thus we will restrict our discussion and analysis only to the linear transformation, i.e. the FrFT and not to the Wigner-Ville distribution.

Based on these observations, in Singh & Datcu [2012b], we proposed a method to analyze SAR images using the FrFT. Later extensive study on this method has been presented in Singh & Datcu [2013]. The FrFT is applicable to any single-look complex (SLC) SAR image. In the following, we will present a basic theoretical background of the FrFT. Later its applicability to the SAR image analysis will be discussed in detail.

5.4.1 Note on the fractional Fourier transform of complex-valued signals

The FrFT was first proposed in mathematical literature in Namias [1980] as a theory of fractional powers for the Fourier transform. Later a formal rigorous study of the FrFT was carried out in Mcbride & Kerr [1987]. The FrFT of a signal $x(t)$ is represented as:

$$X_\alpha(u) = \int_{-\infty}^{\infty} x(t) K_\alpha(t,u) dt, \tag{5.16}$$

where $\alpha = p\pi/2$ denotes the transform angle, with p as a real number called the 'order of the FrFT'. The transformation kernel in the equation 5.18 is defined as follows (Almeida [1994]):

$$K_\alpha(t,u) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}} & e^{j\frac{t^2+u^2}{2}\cot\alpha - jut\csc\alpha}, \\ & \text{if } \alpha \text{ is not a multiple of } \pi \\ \delta(t-u) & \text{if } \alpha \text{ is a multiple of } \pi \\ \delta(t+u) & \text{if } \alpha+\pi \text{ is a multiple of } 2\pi \end{cases} \tag{5.17}$$

The inverse FrFT with transform angle α is the FrFT with the transform angle $-\alpha$, as follows

$$x(t) = \int_{-\infty}^{\infty} X_\alpha(u) K_{-\alpha}(u,t) du, \tag{5.18}$$

Thus, the FrFT can be assumed as the representation of the signal $x(t)$ in terms of a orthonormal basis formed by a set of functions $K_{-\alpha}(u,t)$, with u acting as the parameter for spanning the set of basis functions (Almeida [1994]). This basis function is actually the chirps with a certain frequency modulation rate (FM rate).

It can be inferred from the equation. 5.17 that, for $\alpha = \pi/2$, the FrFT takes the form of a canonical Fourier transform and, for $\alpha = 0$ or a multiple of 2π, it takes the form of an identity-operator. The Fourier transform is based on the decomposition of signals into sinusoids. Whereas, the FrFT can be considered as a decomposition of signals in terms of chirps, with FM rate of the chirp defined by the transform angle α. It can also be mentioned that the FrFT satisfy following basic properties, with ($\mathbf{F}^{\alpha}[\cdot]$) as the FrFT operator (Sun *et al.* [2002]):

1. Linear operator: $\mathbf{F}^{\alpha}[a\ x_1(t) + b\ x_2(t)] = a\ \mathbf{F}^{\alpha}[x_1(t)] + b\ \mathbf{F}^{\alpha}[x_2(t)]$.
2. Identity operator: $\mathbf{F}^{0}[x(t)] = \mathbf{F}^{2\pi}[x(t)] = x(t)$.
3. Consistency with the Fourier transform: $\mathbf{F}^{\pi/2}[x(t)] = X\omega$.
4. Additivity of rotation: $\mathbf{F}^{\alpha+\beta}[x(t)] = \mathbf{F}^{\alpha}[x(t)]\mathbf{F}^{\beta}[x(t)]$.

On the implementation point of view, the slightly modified form presented in Almeida [1994] and Capus & Brown [2003] changes the complex amplitude factor for computational simplicity, by rewriting the FrFT as :

$$X_{\alpha}(u) = \frac{e^{-j(\frac{1}{4}\pi \mathrm{sgn}\alpha - \frac{1}{2}\alpha)}}{(2\pi|\sin(\alpha)|)^{1/2}} e^{\frac{1}{2}ju^2 \cot\alpha} \times \int_{-\infty}^{\infty} e^{-jut\csc\alpha} e^{\frac{1}{2}jt^2\cot\alpha} x(t)dt. \tag{5.19}$$

Here, t and u can interchangeably represent time and frequency domains. The transform output lies between the time and frequency domains, except for the special cases of $\alpha = 0$ and $\alpha = \pi/2$. Based upon the equation 5.19, the FrFT can be realized in a four step process as follows (Almeida [1994], Ozaktas *et al.* [1996]):

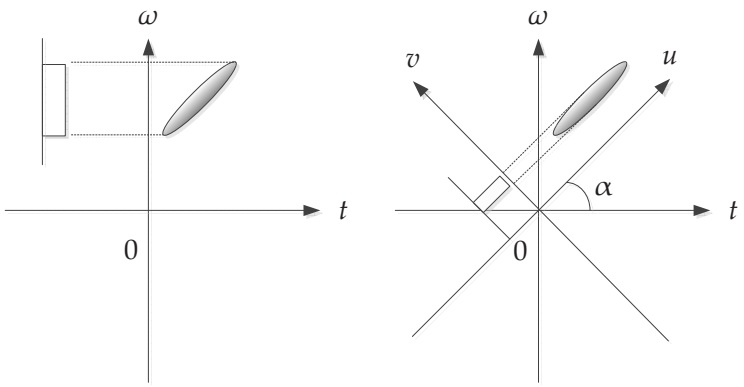
1. Pre-multiplication of input signal by a linear chirp with frequency modulation (FM) rate determined by the transform order.
2. Computation of scaled Fourier Transform with scaling factor of $\sin\alpha$.
3. Post-multiplication by a chirp in transform domain.
4. Another multiplication by a complex amplitude factor.

Using the discrete FrFT formulations presented in Bultheel *et al.* [2002], the N-point FrFT can be computed with a computational complexity of $O(N \log\ N)$ with sufficient accuracy.

The advantages of the FrFT can be understood by considering the case of a linear frequency modulated (LFM) signal, i.e. a chirp signal with some FM rate. Chirp is a typical non-stationary signal, whose instantaneous frequency rises linearly with time with FM rate (Sun *et al.* [2002]). In the canonical Fourier transform (FT) domain, the signal can be regarded as the projection of its time-frequency distribution only onto the frequency axis ω. This leads to spread of the spectrum for a chirp signal, as shown in the figure 5.7a. Whereas, the projection of this chirp signal at a rotated time-frequency domain with suitable angle α will be highly concentrated, as shown in the figure 5.7b. The projection of this signal on other possible planes between time-axis and frequency axis will describe the signature of the signal as a whole.

5.4.2 SAR signal analysis using the fractional Fourier transform

In order to demonstrate the signal energy distribution within the rotated time-frequency plane for fractional orders of the FrFT for SAR signals, the case of a 1-D complex-valued signal extracted from one line in azimuth at a constant range from a SAR image will be considered. We consider to separate examples with signal dimension of 100 pixels, shown in the figures 5.8a



(a) Canonical FT domain (b) The FrFT domain

Figure 5.7: Projections of a linear frequency modulated signal on (a) Fourier transform (FT) domain and (b) the fractional Fourier transform (FrFT) domain. Here, t and ω represents the time and frequency coordinates. The time-frequency coordinate is rotated by an angle α relative to the original coordinates (t, ω) to obtain a set of new coordinates (u, v). This set of new coordinates (u, v) represents the fractional Fourier transform space.

and 5.9a. The thin black line and the thin gray line represent the real and imaginary parts of the signal, respectively. The thick black line is the amplitude of the signal.

The signal in the figure 5.8a constitutes a signal originated from the Doppler shift due to a target moving in the azimuth direction. So its a very good example of a chirp signal with some FM rate. The FM rate of the chirp signal is governed by the velocity of the moving vehicle. Now as shown in the figure 5.8, we will try to find out how this chirp signal is behaving at different planes in the joint time-frequency domain, i.e. at different transform angle in the FrFT domain. For a transform angle of $\alpha = 90°$ (figure 5.8c), which is also the canonical Fourier transform representation of the signal, the signal energy is spread over the whole frequency domain. But, when analyzing the signal using the FrFT at a transform angle of $\alpha = 146.25°$ (figure 5.8d), it can be pointed out that the signal energy is concentrated at a certain location. Figure 5.8e provides an insight into the rotated joint time-frequency domain representation for this signal.

The signal in the figure 5.9a constitutes a signal originated from the azimuth ambiguities visible on the sea surface due to large buildings located at nearby coastal region. The azimuth ambiguity in SAR images occur when the contribution from the backscatter in the antenna footprint to the main lobe of antenna, and the contribution from the backscatter outside of the antenna footprint to the side-lobes of antenna become indistinguishable (Guarnieri [2005]). This happens when azimuth frequency corresponding to both backscatters is equal to a multiple of the sampling frequency or pulse repetition frequency. Now in our case this signal is somehow a complex-valued sinusoidal signal, thus we can see a peak at a particular frequency in the canonical Fourier transform domain (figure 5.9c). Figure 5.9e provides an insight into the rotated joint time-frequency domain representation for this signal.

The basic assumption underlying SAR synthesis is that the backscattering of each contributor is stable regarding the frequency bandwidth and the azimuth viewing angle (Massonnet & Souyris [2008]). The material on ground and directivity of backscattering pattern from various structures produces specific variations in the received signal, and these variations are integrated in the amplitude and phase information in the single-look complex (SLC) SAR images. Such variations are more prominent in the high and very-high-resolution SAR images as the processed

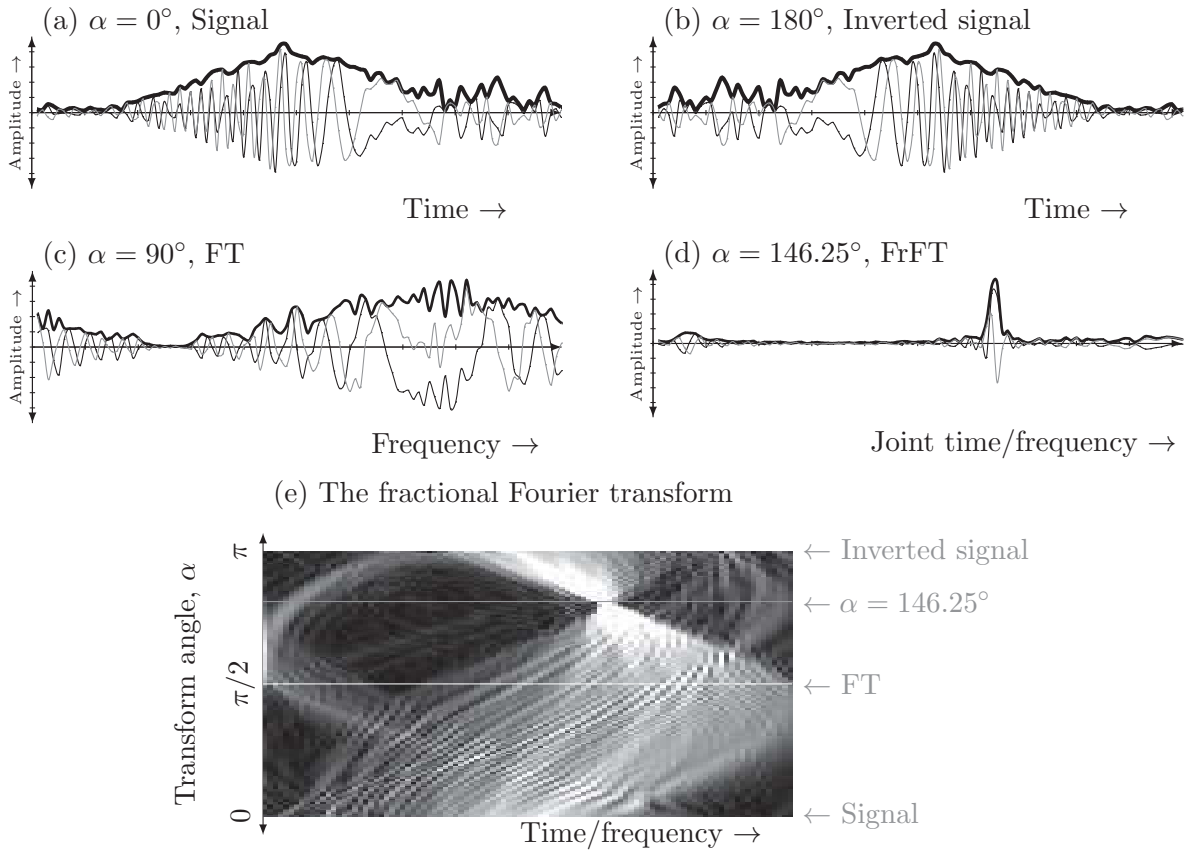


Figure 5.8: The signal energy distribution for some transform angles using the fractional Fourier transform. (a) 1-D signal obtained from a SAR image (one azimuth line of constant range, without Doppler centroid estimation and correction (Madsen [1989], Long *et al.* [2011])), which can also be interpreted as the FrFT at $\alpha = 0°$. (c) the Fourier transform of the signal in (a), which can also be interpreted as the FrFT at $\alpha = 90°$. (d) The FrFT of signal in (a) at $\alpha = 146.25°$, at this transform angle, the energy of the signal is localized at a certain location on the x-axis. (b) The FrFT of the signal at $\alpha = 180°$ is shown only for understanding the FrFT characteristics. (a) to (d) The thick line: normalized magnitude of the signal, the thin black line : real part, the thin gray line: imaginary part. (e) Pictorial representation of signal amplitude variation with respect to transform angle.

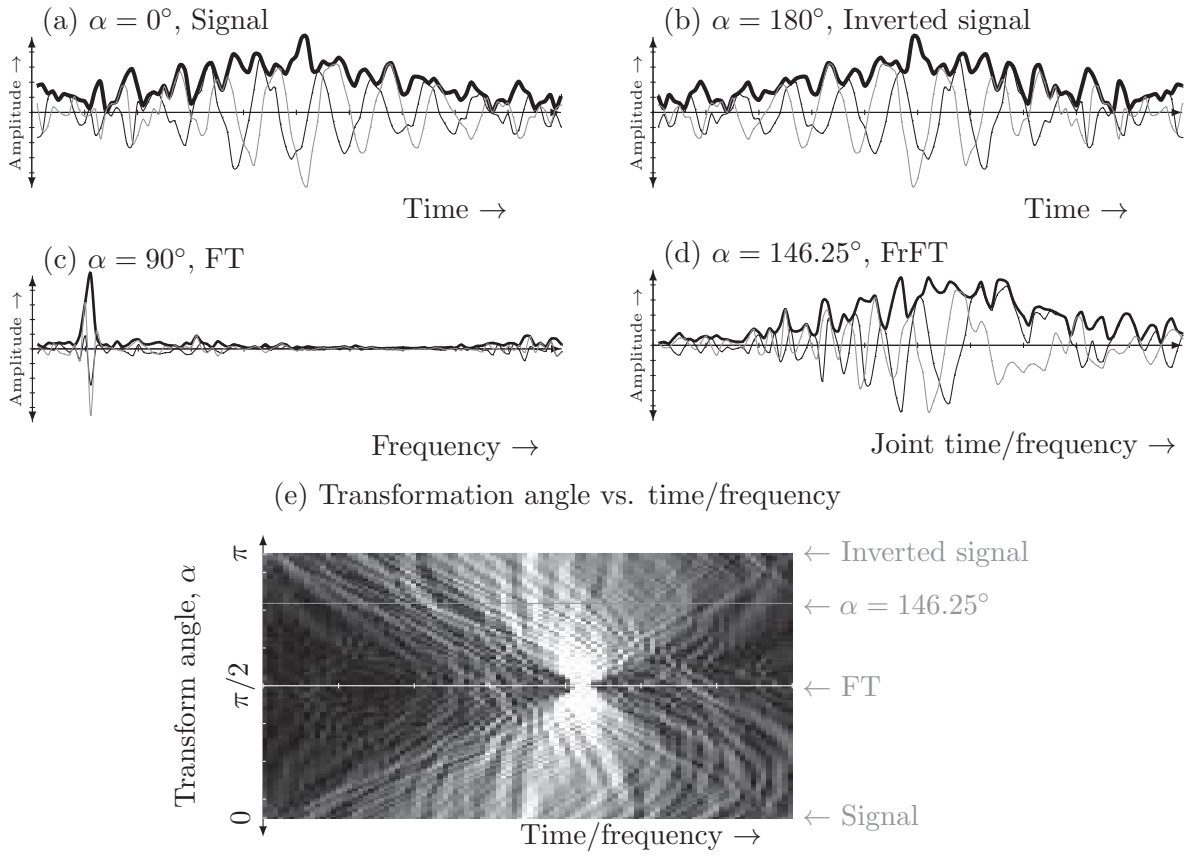


Figure 5.9: The signal energy distribution for some transform angles using the fractional Fourier transform (FrFT). (a) 1-D signal obtained from a SAR image (one azimuth line of constant range, without Doppler centroid estimation and correction (Madsen [1989], Long *et al.* [2011])), which can also be interpreted as the FrFT at $\alpha = 0°$. (c) The Fourier transform of the signal in (a), which can also be interpreted as the FrFT at $\alpha = 90°$. In this particular case, the signal energy is also localized at a certain location on the x-axis for $\alpha = 90°$. (d) The FrFT of the original signal at $\alpha = 146.25°$, here the energy of the signal is distributed over the complete x-axis, unlike the signal energy concentration for the example presented in the figure 5.8. (b) The FrFT of the signal at $\alpha = 180°$ is shown only for understanding the FrFT characteristics. (a) to (d) The thick line: normalized magnitude of the signal, the thin black line : real part, the thin gray line: imaginary part. (e) Pictorial representation of signal amplitude variation with respect to transform angle.

bandwidths are larger (Spigai *et al.* [2011]). The projections of complex-valued backscattered signal on the rotated time-frequency domain will discover these variations and will provide hidden information about the backscattering phenomenon occurring on ground.

In order to find a compact descriptor for SAR signals, we exploit the representation of a signal at various transform angles to demonstrate signal characteristics. We can generate a signal descriptor from projections of signals on various time-frequency planes. We extend this assumption to 2-D signals for the SAR image patch case. The 2-D FrFT can be achieved by two 1-D separable transforms on x and y domains. Apart from discovering the backscattering information in various planes of the joint time-frequency domain, the statistical characterization is simpler in case of the FrFT coefficients as compared to original image patch. The FrFT algorithm behavior is shown in the figure 5.10 with example patches containing various objects. It can be seen in the figure 5.10a, the pdf of the original patch is multimodal (shaded-gray), which is transformed into unimodal at various transform angles in the FrFT space. For the image patch in the figure 5.10b, the pdf for original image patch is heavy-tailed (shaded-gray), which is transformed at various transform angles in the FrFT space into pdfs which are easier to analyze.

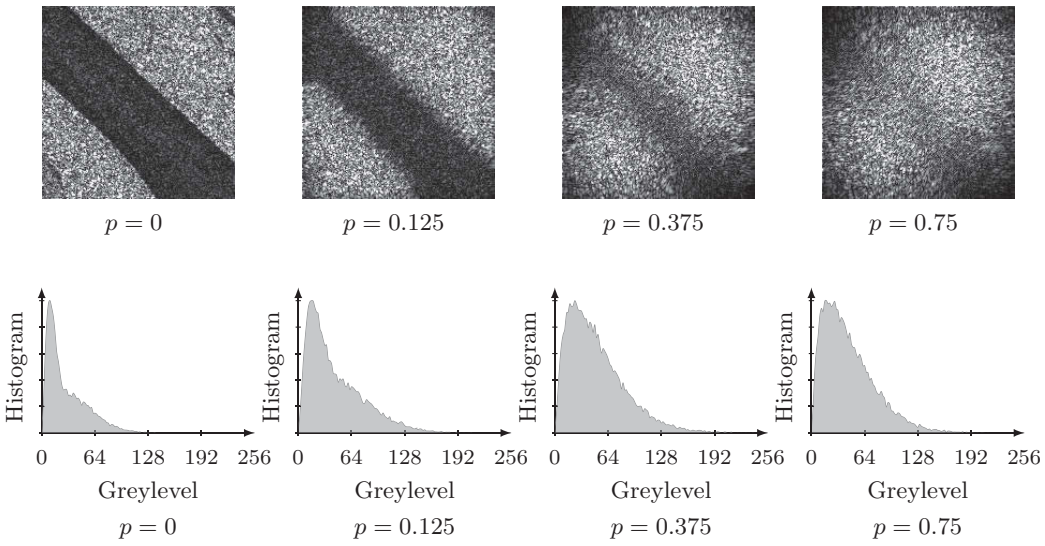
5.4.3 Note on 2-D separable fractional Fourier transform

Being a linear transform, the FrFT can be applied to x domain first and on this output FrFT can be applied to y domain. The same approach is followed while computing the canonical Fourier transform of 2-D signals, i.e. images. So, for an image $I(x, y)$, the 2-D FrFT can be computed by applying the transform first to the rows and then to the columns. However the sequence of the transform can be reversed (i.e. first to the columns and then to the rows), without changing the results as the FrFT is a linear transform. We demonstrate this concern by computing the FrFT for an image-patch using both the orders, as shown in the figure 5.11.

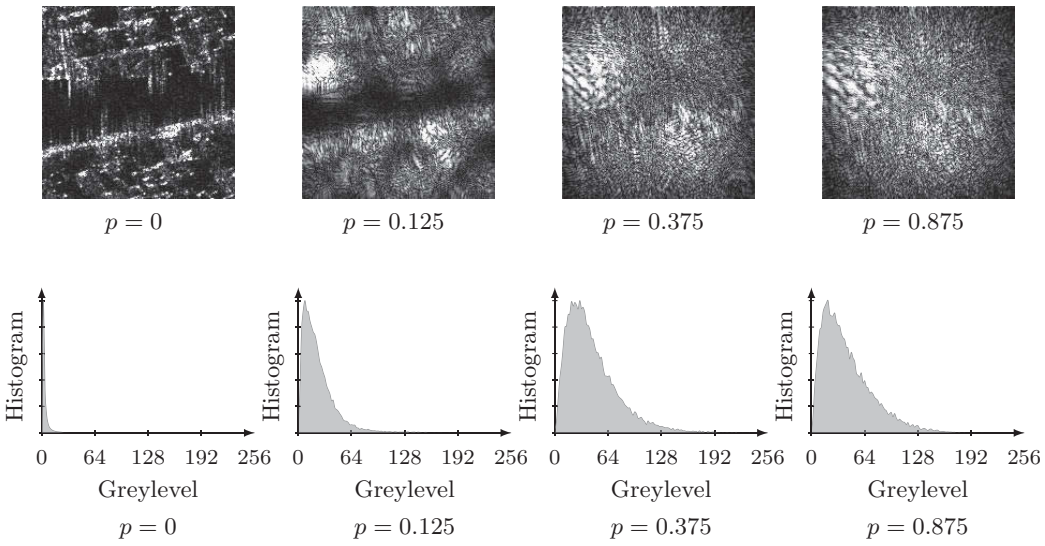
Some structures and objects in the SAR images have peculiar signal response when seen in the fractional Fourier space. In the previous section, we demonstrated this with the help of 1-D signals extraction from SLC SAR images (figures 5.8 and 5.9). Now we will present similar observations on the 2-D signals, i.e. on the SLC SAR images.

Figure 5.12 at $p = 0$ shows the 2-D azimuth ambiguities on the sea surface from large building blocks on the nearby coastal region. The FrFT at certain transform angles with respective real part (top row), imaginary part (middle row) and amplitude (bottom row) are presented. The signatures of these signals in amplitude images may be perceived as ship like structure and confused with ships in image information extraction. With the visualization of various projections in fractional Fourier domain it is possible to suggest the true nature of target, as can be seen in the figure 5.12.

Figure 5.13 at $p = 0$ shows an SAR image patch consisting of buoys on the sea surface. The FrFT at certain transform angles with respective real part (top row), imaginary part (middle row) and amplitude (bottom row) are presented. It has been pointed out earlier that the FrFT can be perceived as the decomposition of signal in terms of chirps. This phenomenon can be observed in the real and imaginary part of the various transform orders for this target. The signal at various transform orders is spread in terms of chirps with FM rate of the chirp relative to the transform angle.



(a) A SAR image patch (shown at $p = 0$) with multimodal density and its fractional Fourier transform at selected transform orders along with corresponding densities. It is evident from analyzing the histogram of the amplitude of the fractional Fourier transform coefficients, that the statistical response turns into unimodal from multimodal of the original SAR image. This will make the statistical analysis of the coefficients easier.



(b) A SAR image patch (shown at $p = 0$) with heavy-tailed density and its fractional Fourier transform at selected transform orders along with corresponding densities. It is evident from analyzing the histogram of the amplitude of the fractional Fourier transform coefficients, that the statistical response is not any more heavy-tailed as for the original SAR image. This will make the statistical analysis of the coefficients easier.

Figure 5.10: The amplitude of the FrFT coefficients of SAR image patches (shown at $p = 0$) with (a) multimodal density behavior and (b) heavy-tailed density behavior for the selected transform orders. Here p is the transform order of the FrFT with $\alpha = p\pi/2$ denoting the transform angles along with normalized histogram of the corresponding amplitude FrFT coefficients.

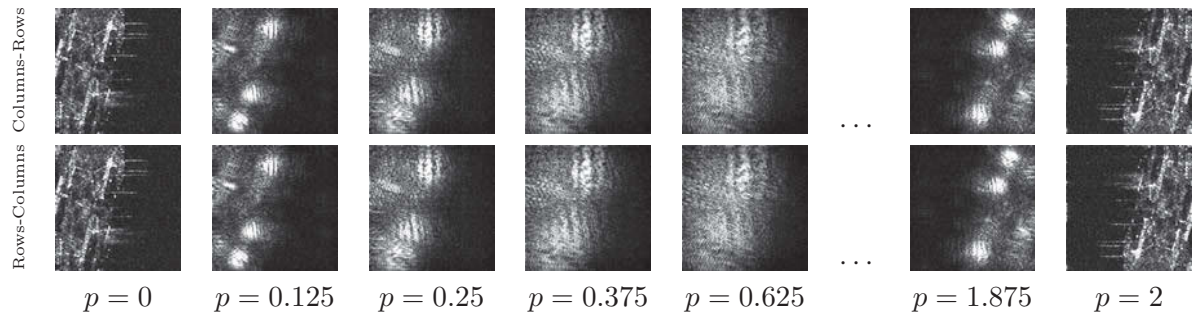


Figure 5.11: Computation of the 2-D fractional Fourier transform of a SAR image patch (shown at $p = 0$) at transform order step $\Delta p = 0.125$. Top : first columns, then rows. Bottom : first rows, then columns. Where p is the transform order of the FrFT with $\alpha = p\pi/2$ denoting the transform angle.

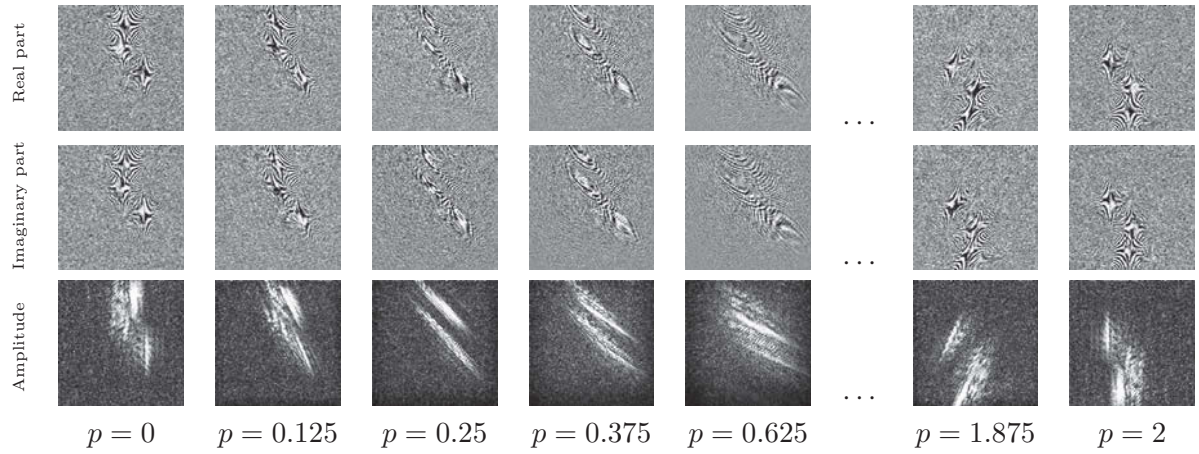


Figure 5.12: (Top) Real part, (Middle) imaginary part and (Bottom) amplitude-envelope of the SAR image patch (shown at $p = 0$) at certain transform orders. Here p is the transform order of the FrFT with $\alpha = p\pi/2$ denoting the transform angle.

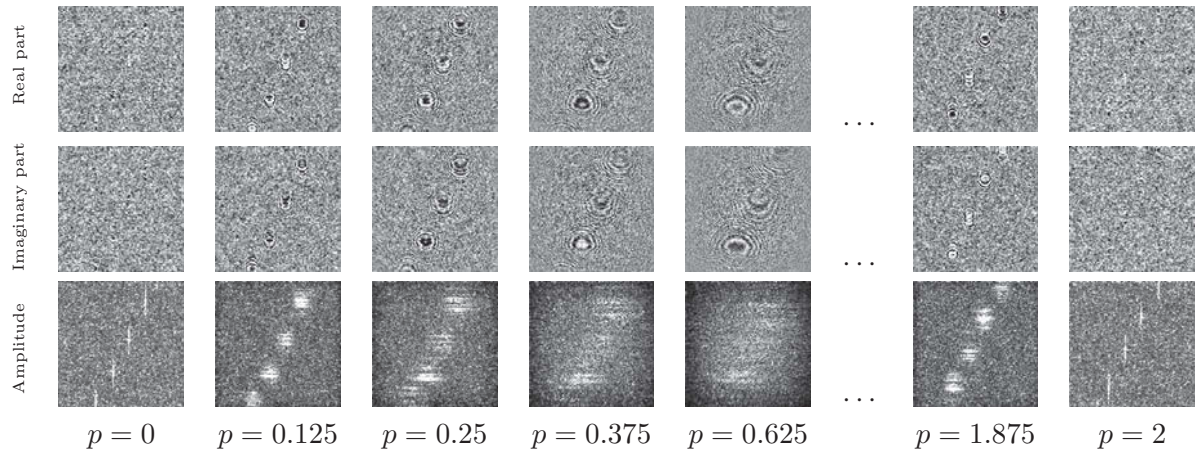


Figure 5.13: (Top) Real part, (Middle) imaginary part and (Bottom) amplitude-envelope of the SAR image patch (shown at $p = 0$) at certain transform orders. Here p is the transform order of the FrFT with $\alpha = p\pi/2$ denoting the transform angle.

5.4.4 Feature descriptors based on the fractional Fourier transform

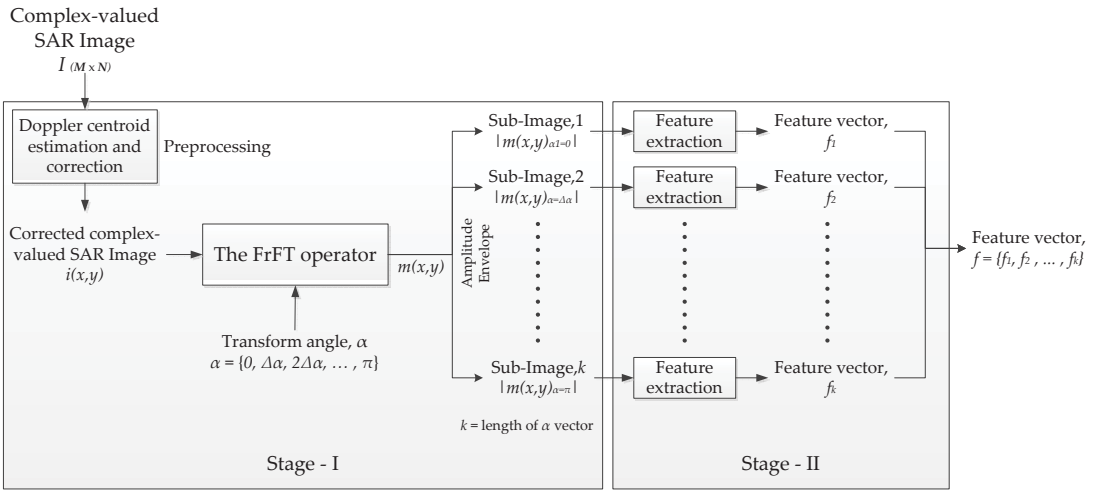


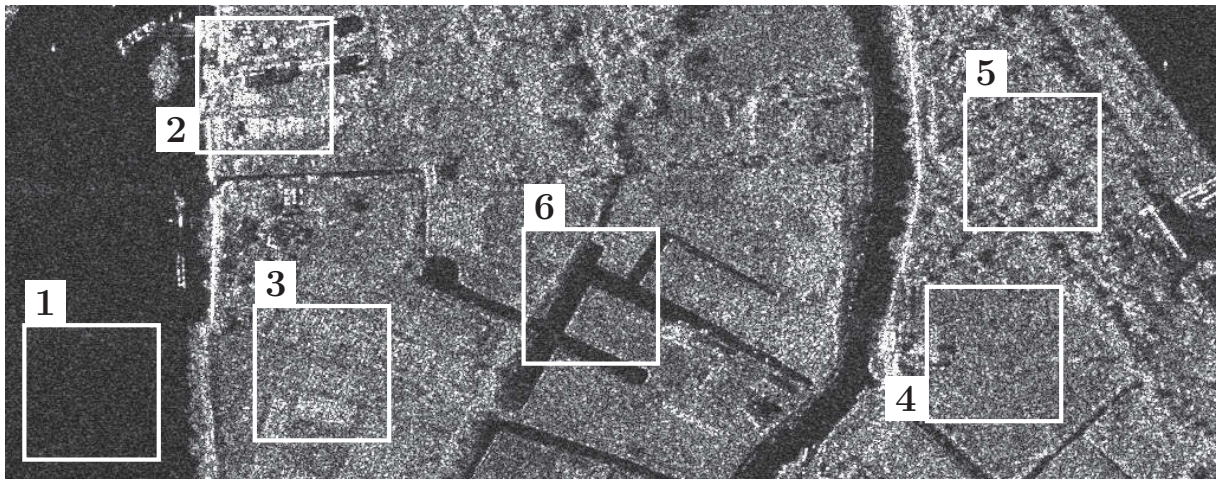
Figure 5.14: Architecture of the sub-images based FrFT nonparametric analysis of images for information extraction.

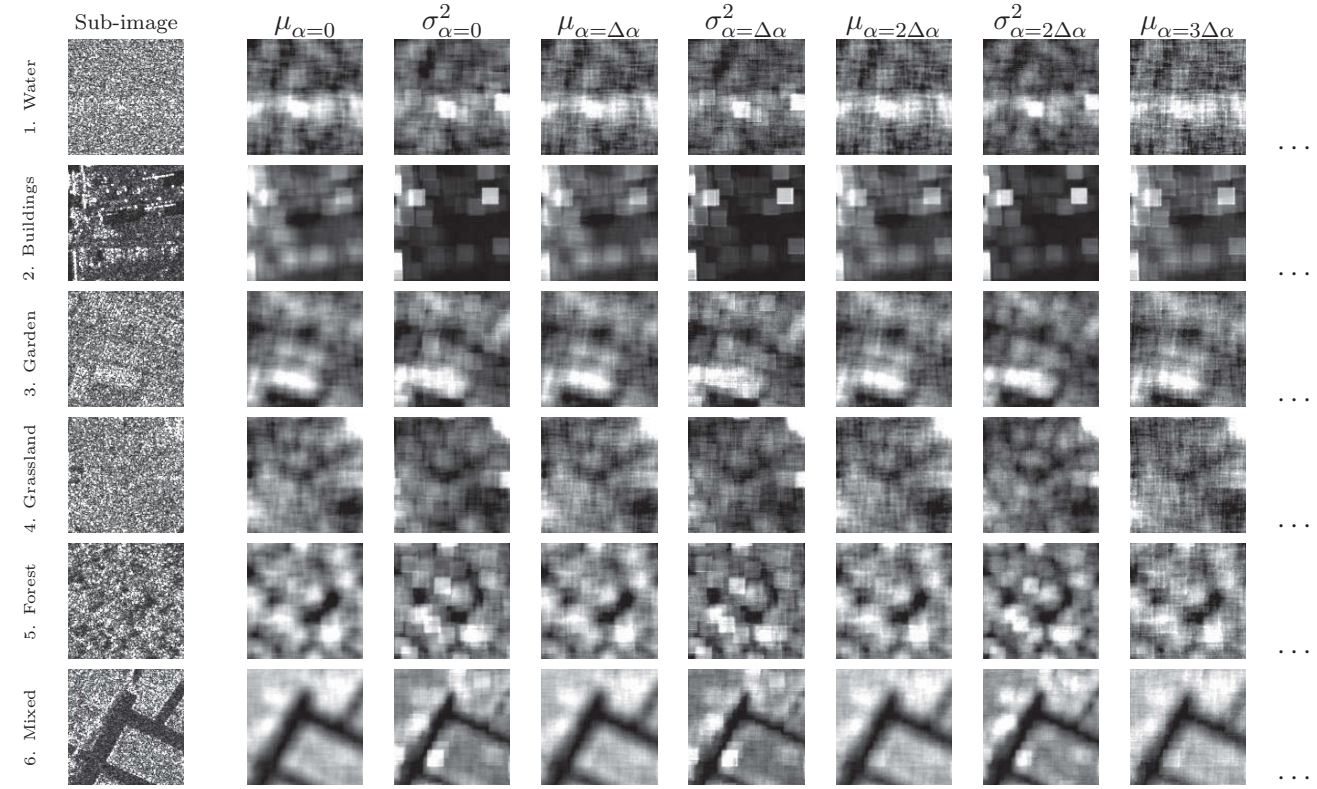
In reference to the figure 5.1, the architecture of generating a FrFT-based feature descriptor is shown in the figure 5.14. Input is a complex-valued SAR image. As a preprocessing step, the Doppler centroids in range and azimuth estimated and correcting with the use of appropriate algorithms (Madsen [1989], Long *et al.* [2011]). The projections of this corrected complex-valued SAR images on rotated joint time-frequency planes are computed using the methodology presented by the equation 5.19 at certain transform angles. This provides us complex-valued sub-images $\mid m(x,y) \mid$ as shown in stage-I in the figure 5.14, on the amplitude $\mid m(x,y) \mid$ of which the feature descriptors are computed in stage-II in the figure 5.14. In the classical study by Julesz in Julesz [1981], it has been demonstrated that only linear mechanisms are not sufficient to explain how human visual system perceive the texture. These principles has been further asserted in Malik & Perona [1990], as well as in Caelli [1985]. This was the reason to compute first the amplitude of the complex-valued sub-bands, over which simple statistical measure are computed to generate a feature descriptor. The obvious choice will be to compute the moments.

In Singh & Datcu [2012b], we proposed the use of first two moments (mean, μ, and variance, σ^2) of the amplitude of the FrFT coefficients for the patch-oriented image categorization. Considering a suitable interval of $\Delta\alpha = 0.125$, with $\{\alpha = 0, \Delta\alpha, 2\Delta\alpha, 3\Delta\alpha, \cdots, \pi\}$, the feature descriptor in our analysis for patch-oriented image categorization also includes the third moment (skewness, γ), which has the following form, of length 51 at $\Delta\alpha = 0.125$:

$$\bar{f}_{\mathrm{FrFT}} = \{\mu_0, \sigma_0^2, \gamma_0, \mu_{\Delta\alpha}, \sigma_{\Delta\alpha}^2, \gamma_{\Delta\alpha}, \cdots, \mu_\pi, \sigma_\pi^2, \gamma_\pi\}. \tag{5.20}$$

For pixel-based classification we exclude the third moment (skewness, γ) in the feature descriptor, consequence of which the length of the feature descriptor reduces to 34 at $\Delta\alpha = 0.125$. An example of computed feature descriptor $\bar{f}_{\mathrm{FrFT}}$ based on the moments of the amplitude of the FrFT coefficients and visual inspection for texture separability is presented in the figure 5.15.





Sub-image	$\mu_{\alpha=0}$	$\sigma^2_{\alpha=0}$	$\mu_{\alpha=\Delta\alpha}$	$\sigma^2_{\alpha=\Delta\alpha}$	$\mu_{\alpha=2\Delta\alpha}$	$\sigma^2_{\alpha=2\Delta\alpha}$	$\mu_{\alpha=3\Delta\alpha}$	$\cdots$
1.	3.424e+1	3.787e+2	3.676e+1	2.203e+2	3.237e+1	4.121e+2	3.038e+1	$\cdots$
2.	9.364e+1	4.421e+3	9.752e+1	3.546e+3	8.798e+1	4.767e+3	7.860e+1	$\cdots$
3.	1.595e+2	7.321e+3	1.665e+2	5.306e+3	1.468e+2	9.227e+3	1.345e+2	$\cdots$
4.	1.052e+2	3.021e+3	1.105e+2	2.210e+3	9.723e+1	3.767e+3	8.721e+1	$\cdots$
5.	1.615e+2	1.203e+4	1.744e+2	7.656e+3	1.538e+2	1.210e+4	1.446e+2	$\cdots$
6.	6.696e+1	3.476e+3	6.941e+1	2.842e+3	6.557e+1	3.101e+3	5.912e+1	$\cdots$

Figure 5.15: The fractional Fourier transform based feature descriptor. The value for $\Delta\alpha = \pi/16$. Similarities can be seen in the texture separability using the FrFT-based feature descriptor with the non-linear spectral feature descriptor computed in the Fourier domain. But we see enhancement on a larger scale. E.g. the rough sea on the lower part in sub-image 1 is very well distinguishable. Similar observation can be made for the specific texture for the lower-left region in sub-image 3. The computation of the fractional Fourier transform coefficients is rather computationally expensive. It takes around 30 minutes to process one sub-image at presented settings, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM.

5.5 Gabor filter banks

The use of the FrFT is limited to the complex-valued signals, whereas another joint time-frequency analysis method, the Gabor expansion is applicable to the real-valued signals only. Gabor expansion introduced by a Hungarian-born British physicist, Dennis Gabor, in 1946, particularly suitable for analysis of signals with sharp changes but which stay for a small duration of time. It suggests expanding a signal into a set of functions that are concentrated in both time and frequency (Gabor [1946]). This set of functions, also called the Gabor coefficients, plays the main role in analysis of signal's local properties. Gabor expansion can be written mathematically as follows (Qian & Chen [1993]):

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_{m,n} h_{m,n}(t), \text{with} \quad h_{m,n}(t) = h(t - mT) e^{jn\omega t}. \tag{5.21}$$

Where, $G_{m,n}$ are called the Gabor coefficients. The set of elementary functions $\{h_{m,n}(t)\}$ consists of a time-frequency shifted function $h(t)$. In practice to describe the texture of an image $i(x, y)$, a filter structure is employed where image is filtered using the bandpass filters, also called Gabor elementary functions $h(x, y)$, to extract a specific frequency band from an image. Usually a convolution operation $(*)$ of GEF's, $h(x, y)$, with the original image, $i(x, y)$, is employed to obtain a Gabor filtered image, $m(x, y)$, as follows:

$$m(x, y) = |i(x, y) * h(x, y)|. \tag{5.22}$$

5.5.1 Description of the Gabor elementary functions

Gabor coefficients or Gabor elementary functions (GEF's) are not unique and can be described in many different ways. GEF's are sometimes also referred as Gabor wavelets, which are nothing but the Gaussian modulated by a complex sinusoid as follows (Super & Bovik [1991], Bovik *et al.* [1990]):

$$h(x, y) = g(x', y') e^{j2\pi(Ux+Vy)} \tag{5.23}$$

Here, $(x', y') = (x \cos\phi + j \sin\phi, -x \sin\phi + y \cos\phi)$ are the rotated rectilinear coordinates in spatial domain, and (U, V) are the particular 2-D frequency in the complex-sinusoid. The 2-D Gaussian function $g(x, y)$ with its corresponding Fourier transform $G(u, v)$ is give as (Dunn *et al.* [1994]):

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_u^2} + \frac{y^2}{\sigma_y^2}\right)\right], \tag{5.24}$$

$$G(u, v) = \exp\{-2\pi^2 \left[(\sigma_x [u - U]')^2 + (\sigma_y [v - V]')^2\right]\}. \tag{5.25}$$

Here, $[(u - U)', (v - V)'] = [(u - U)\cos\theta + (v - V)\sin\theta, -(u - U)\sin\theta + (v - V)\cos\theta]$ are the shifted and rotated rectilinear coordinates in the frequency-domain, with $\sigma_u = 1/2\pi\sigma_x, \sigma_v = 1/2\pi\sigma_y$ characterizing the spatial extent and bandwidth of $h(x, y)$. Filter parameters (σ_u and σ_v) can be computed in several ways. The Gabor wavelets are non-orthogonal in nature unless they are poorly concentrated in time or in frequency domain. While keeping a good time and frequency domain concentration, non-orthogonality is allowed at the cost of redundant information in the filtered image. In order to reduce this redundancy, following formulation to

compute σ_u and σ_v has been proposed in Manjunath & Ma [1996]:

$$\begin{aligned}\sigma_u &= \frac{((U_h/U_l)^{\frac{1}{\mathbf{S}-1}}-1)U_h}{((U_h/U_l)^{\frac{1}{\mathbf{S}-1}}+1)\sqrt{2\log 2}},\\ \sigma_v &= \tan\left(\frac{\pi}{2\mathbf{O}}\right)\left[U_h - 2\log\left(\frac{2\sigma_u^2}{U_h}\right)\right]\left[2\log 2 - \frac{(2\log 2)^2\sigma_u^2}{U_h^2}\right]^{-\frac{1}{2}}.\end{aligned} \tag{5.26}$$

With, U_h and U_l are the upper and lower center frequencies of interest in the Gabor kernel, $\mathbf{S}$ is the number of stages and $\mathbf{O}$ is the number of orientations used in the multi-scale/ multi-resolution Gabor decomposition. The choice of Gabor coefficients is not restrictive but the foundations for wide usage of Gabor filter banks for texture description were mainly laid in Manjunath & Ma [1996] with the proposal of a homogeneous texture (HT) descriptor, and later advocated in diverse fields studies (Kandaswamy *et al.* [2005], Riaz *et al.* [2012]). we also followed the same formulation mentioned in the equation 5.26, with $U_h = 0.4$, $U_l = 0.05$.

In the figure 5.16, we illustrate the real and imaginary parts of the Gabor kernel at Stage $\mathbf{S} = 1$, and six different orientations $\mathbf{O} = 1, 2, \cdots, 6$. The SAR image under consideration, or a analyzing window in SAR image is convolved with this complex-valued Gabor filter bank to obtain a Gabor filtered image.

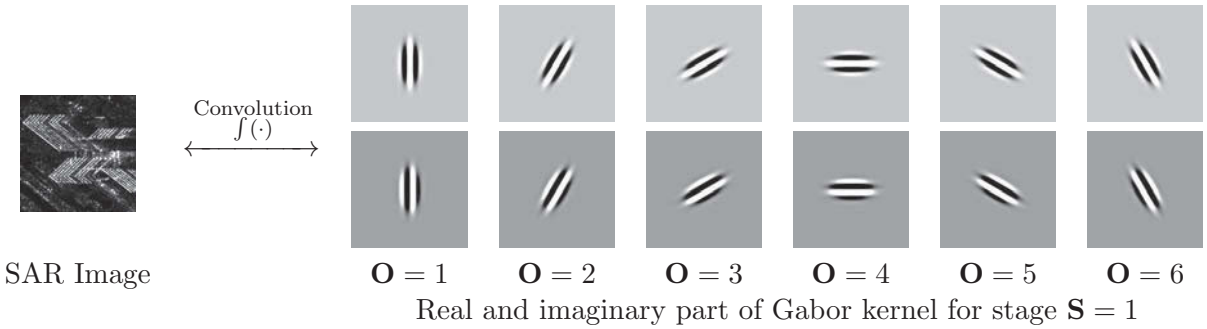


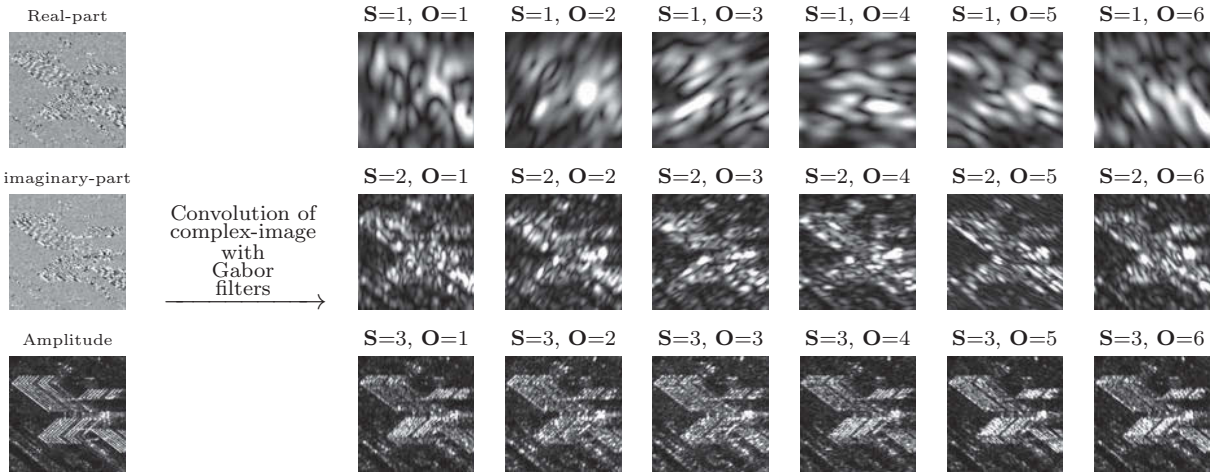
Figure 5.16: SAR image is convolved with the Gabor filter banks to get a Gabor filter image. Shown are (top) the real part and (bottom) the imaginary part of the Gabor kernel for a particular stage $\mathbf{S} = 1$, and six different orientations $\mathbf{O} = 1, 2, \cdots, 6$, using the variance formulation presented in the equation 5.26. Here $\int(\cdot)$ depicts the convolution operator.

Convolution of SAR image or analyzing window in SAR image with complex-valued Gabor will provide us a complex-valued Gabor filtered output. The amplitude of this filtered output is used to generate a feature descriptor.

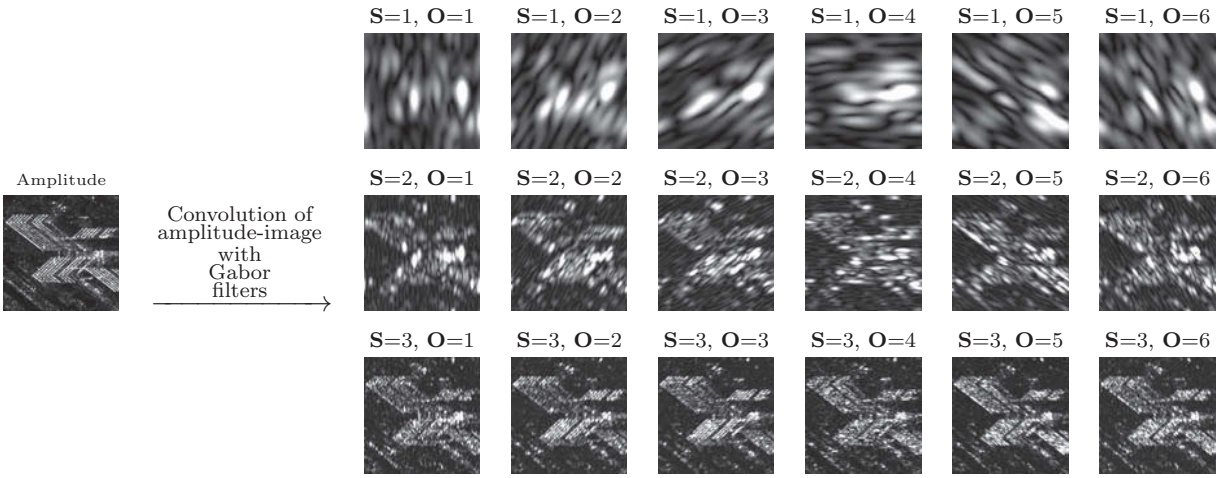
5.5.2 Note on Gabor filter banks applied to complex-valued SAR images

We have seen that the most essential property of Gabor coefficients for characterizing the texture in images is its bandpass filter property, which has also been demonstrated in Farrokhnia & Jain [1991]. This bandpass filter like operation is also the underlying mechanism in human visual systems, which we try to imitate in the Gabor expansion of signals to understand the texture. Now we must emphasize here that Gabor filter banks applied to the complex-valued SAR images will loose this true bandpass filter property. The reason is that carrying out a bandpass filtering with Gabor like operation for complex-valued SAR data will lead to a kind of multiple sub-look decomposition (Spigai *et al.* [2011], Singh & Datcu [2012c]), which is not capturing the textural information. Thus the Gabor filters applied to complex-valued SAR images provides a very

poor accuracy in the classification. So it must be noted that the Gabor filters in our case have been used for the amplitude SAR images. This concern is demonstrated in the figure 5.17, by convolving the complex-valued and amplitude SAR image with Gabor filter banks.



(a) Gabor filter applied to complex-valued SAR image.



(b) Gabor filter applied to amplitude SAR image.

Figure 5.17: SAR image is convolved with the Gabor filter banks to get a Gabor filter image. Shown are the amplitudes of the Gabor filtered images using 3 stages and 6 orientations, using the variance formulation presented in the equation 5.26. The amplitudes of the Gabor filtered images corresponds to the sub-images $\mid m(x, y) \mid$ in the stage-I of the figure 5.19. Feature descriptors are extracted from these sub-images in stage-II.

In the figure 5.17a, we apply the Gabor filter bank shown in the figure 5.16 to complex-valued SAR image, and in the figure 5.17b, we apply this filter bank to amplitude SAR image. It can be seen that whereas, the information about the texture is preserved in the 5.17b by forming blob-like directional textures in accordance to the directivity of the corresponding filter bank used, in figure 5.17a we see various multiple sub-looks filtered with rotated complex-valued Gaussian. However this phenomena is not significantly visible at stage $\mathbf{S} = 1$, but can be seen at lower stages $\mathbf{S} = 2$ and 3. Some selected enlarged filtered images for both cases are shown in the figure 5.18. Thus the application of the Gabor filter banks to the complex-valued SAR

images will not be accurately capturing the textural information. This will lead to an inaccurate feature descriptor and in turn a reduced performance.

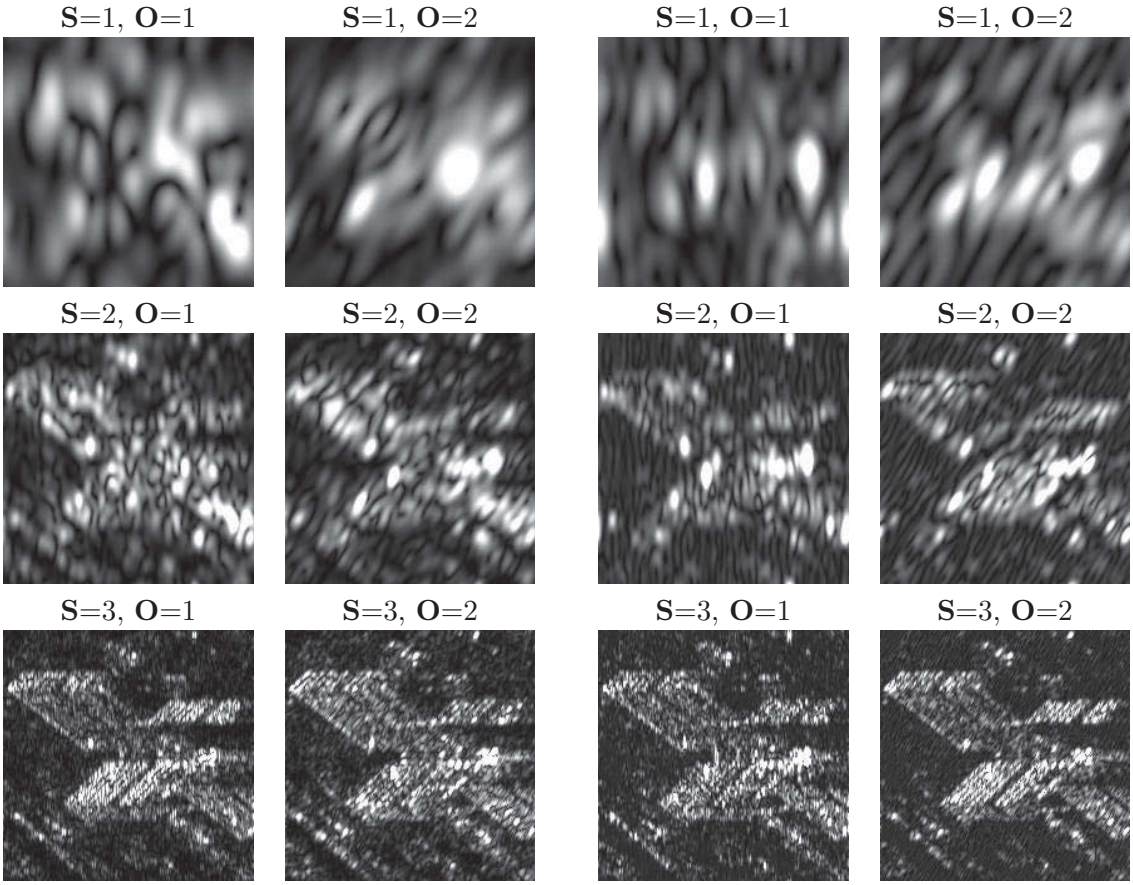


Figure 5.18: Few selected enlarged Gabor filtered SAR image patches from figure 5.17. It is evident from the filtered images generated from the amplitude images shows the oriented sub-bands. Whereas filtered images generated from the complex-valued SAR images are a kind of sub-looks extracted from the complex-valued spectrum.

5.5.3 Feature descriptors for SAR images based on Gabor filter banks

Even for the Gabor filters applied to amplitude SAR images, we will obtain complex-valued filtered images. Again as it has been commented in Julesz [1981], that only linear mechanisms are not sufficient to explain how human visual system perceive the texture, to further incorporate non-linearity in the Gabor based approach, one idea is to work on the amplitude of the complex-valued sub-images. In Bovik *et al.* [1990], as well as in Manjunath & Ma [1996], it has been studied extensively that amplitude of the complex-valued sub-images can be used to characterize the texture of images.

Now the obvious choice for a feature descriptor over the amplitude of the filtered images are again the simple statistical measure, such as mean μ and variance σ^2 etc. Thus, we also compute the moments of the amplitude of the complex-valued Gabor filtered output at different scales and orientations, which later constitutes the feature descriptor. The same hypothesis

motivated us to propose the feature descriptor based on the moments of the amplitude of the complex-valued FrFT coefficient, which has been already discussed in section 5.4.

The invariance properties of Gabor filters along with the restrictions to the use of Gabor filters for feature extraction has been demonstrated in Kamarainen *et al.* [2006]. Even though a lot of study have been carried out since the Gabor expansion with the proposal of some recent feature descriptor based on Gabor filters such as auto-correlation Gabor features (Riaz *et al.* [2012]) etc., the use of moments of the Gabor filtered images are still widely accepted tool for texture synthesis in multimedia as well as remote sensing images. The homogeneous texture descriptor under the visual texture descriptor in the moving picture experts group (MPEG)-7 standard (Sikora [2001]) also rely on this feature descriptor, thus can be considered as a benchmarking tool for comparing alternative methodologies. In reference to the figure 5.1, the architecture of generating Gabor-based feature descriptor is shown in figure 5.19.

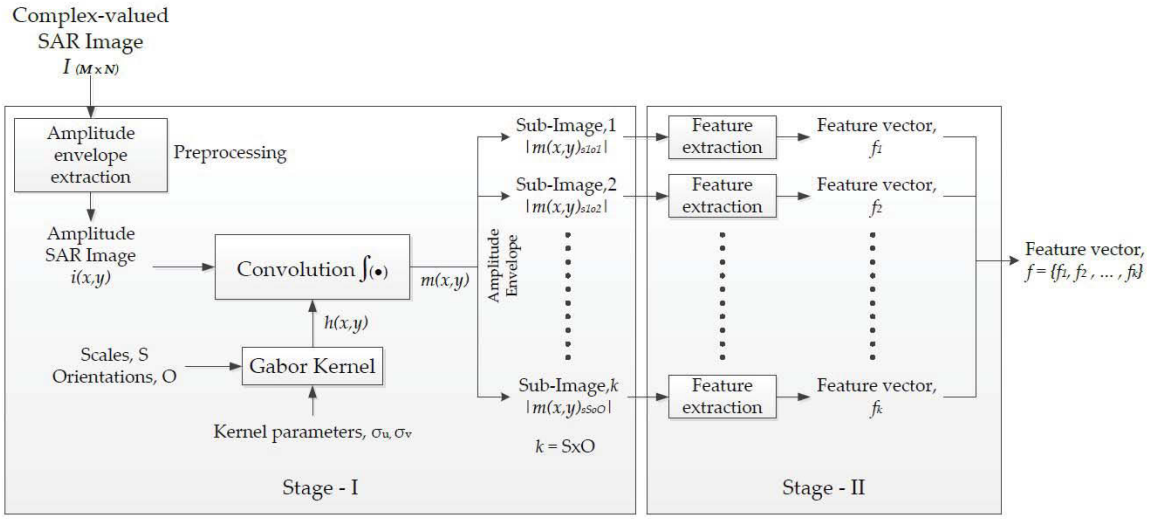
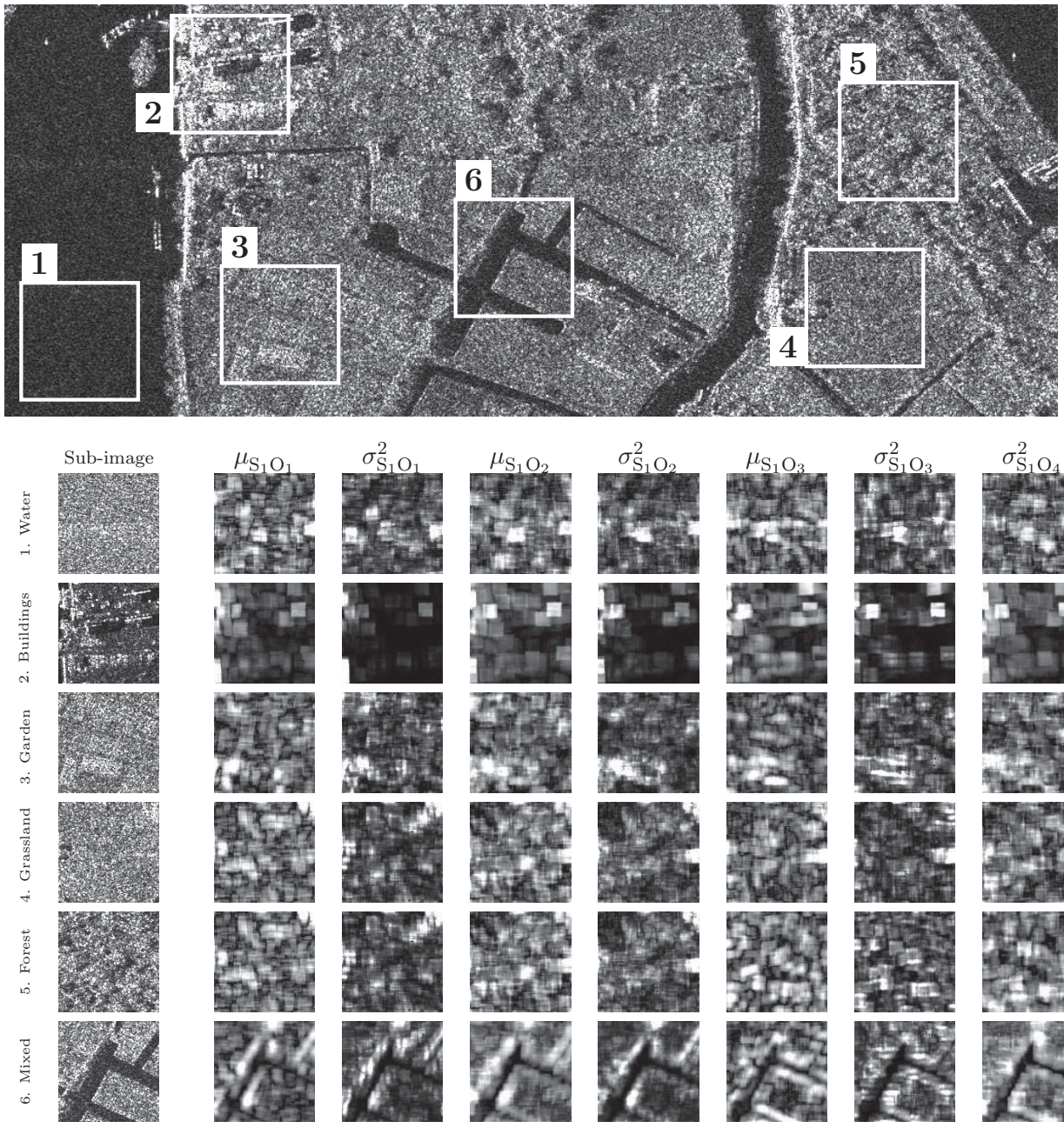


Figure 5.19: Architecture of the sub-images based Gabor nonparametric analysis of images for information extraction.

In our study, we have used 3 stages, and 4 orientations, with $U_h = 0.4$, $U_l = 0.05$ in Gabor filter bank formation. The feature descriptor in our analysis for patch-oriented image categorization also includes the third moment (skewness, γ). Thus, for 3 stages, and 4 orientation, we obtain a feature descriptor of length 36 as follows:

$$\begin{aligned}\bar{f}_{\text{Gabor}} = \{&\mu_{1,1}, \sigma^2_{1,1}, \gamma_{1,1}, \mu_{1,2}, \sigma^2_{1,2}, \gamma_{1,2}, \mu_{1,3}, \sigma^2_{1,3}, \gamma_{1,3}, \mu_{1,4}, \sigma^2_{1,4}, \gamma_{1,4},\\ &\ldots \mu_{3,1}, \sigma^2_{3,1}, \gamma_{3,1}, \mu_{3,2}, \sigma^2_{3,2}, \gamma_{3,2}, \mu_{3,3}, \sigma^2_{3,3}, \gamma_{3,3}, \mu_{3,4}, \sigma^2_{3,4}, \gamma_{3,4}\}.\end{aligned} \tag{5.27}$$

With, $\mu_{\mathbf{S},\mathbf{O}}$, $\sigma^2_{\mathbf{S},\mathbf{O}}$ and $\gamma_{\mathbf{S},\mathbf{O}}$ are the first three moments (mean, variance and skewness, respectively) of Gabor filtered image at stage-**S**, orientation-**O**. For pixel-based classification we exclude the third moment (skewness, γ) in the feature descriptor, consequence of which for 3 stages, and 4 orientation, we obtain a feature descriptor of reduced length 24. An example of computed feature descriptor $\bar{f}_{\text{Gabor}}$ based on the moments of the amplitude-envelope of the Gabor filtered images and visual inspection for texture separability is presented in the figure 5.20.



Sub-image	$\mu_{S_1O_1}$	$\sigma^2_{S_1O_1}$	$\mu_{S_1O_2}$	$\sigma^2_{S_1O_2}$	$\mu_{S_1O_3}$	$\sigma^2_{S_1O_3}$	$\mu_{S_1O_4}$	$\sigma^2_{S_1O_4}$	$\mu_{S_2O_1}$	$\sigma^2_{S_2O_1}$	$\cdots$
1.	1.27	0.36	4.78	5.65	11.33	0.39	6.35	7.18	4.19	8.62	$\cdots$
2.	22.08	5.48	26.98	87.98	72.47	10.30	27.70	123.14	13.73	89.03	$\cdots$
3.	20.30	5.39	32.87	131.63	32.04	10.95	16.03	54.82	15.28	116.05	$\cdots$
4.	11.50	1.62	20.67	25.25	15.71	1.26	13.41	28.54	12.77	76.86	$\cdots$
5.	14.59	6.84	49.31	485.50	90.63	19.63	34.86	143.72	24.38	271.04	$\cdots$
6.	37.94	5.45	23.70	111.68	54.81	7.03	30.21	245.72	11.46	104.71	$\cdots$

Figure 5.20: Gabor filter banks based feature descriptor. Analyzing window used is 16×16 pixels. Only 2 stages and 6 orientations have been used in the Gabor filter bank. Its evident that at lower analyzing window, the performance in texture / category separability is relatively poor for Gabor-based methods. However, its a very computationally inexpensive method. It takes less than a minute to process one sub-image at presented settings, on a standard laptop with Intel Core processor (2.40 GHz) with 8 GB RAM.

5.6 Summary

The following points have been discussed in this chapter.

- The nonparametric information extraction approaches are the methods where no specific model for a given data is postulated, which provides us more flexibility in the spatial content understanding. Similar to the parametric information extraction approaches, the parametric information extraction approaches for spatial content understanding can also be further classified into approaches applied to images in spatial domain and approaches applied to images within a transformation domain. We have considered only the sub-image based approaches in our analysis.

- One of the very popular nonparametric approaches applied to images in spatial domain is the statistical texture features computed on the gray level co-occurrence matrices. Theoretical background and considered textural features have been presented for the sake of completeness.

- The other important nonparametric approaches applied to the images within a transformation domain generally are the time-frequency approaches. In time-frequency analysis the processing on the spectrum of the SAR images is carried out to generate a descriptor for the SAR images. The simplest approach would be to carry out the non-linear features computed on the spectrum. We proposed few non-linear features computed on the spectrum motivated from the timbrel texture features in music genre.

- The other time-frequency approaches are generally multi-scale approaches, such as Gabor filter banks. The Gabor filter bank has been extensively applied to the amplitude SAR images for information extraction. We have discussed the reasoning of non-applicability of Gabor filter banks like multi-scale approaches to the complex-valued SAR images.

- Limitations of the Gabor filter bank like approaches for complex-valued SAR images motivated us to propose the use of fractional Fourier transform, which is actually a true multi-scale approach for complex-valued SAR images where the scaling is carried out in the phase. This will aid us to take advantage of the phase information for SAR information extraction, which has been earlier demonstrated and discussed in the MSLD approach.

- Preliminary results on information extraction for above considered methods have been presented.

As a summary of the Part-II, the main focus of the discussion in the earlier chapter, chapter 4, was on the parametric approaches and in the current chapter on the nonparametric approaches. In next part, part-III, we will analyze their respective performances in context to the methodological framework proposed in chapter 3. In the following, the results, discussions and conclusion will be presented in detail.

Part III

Results and Conclusions

Chapter 6

Illustrations and case-study

The objective of this chapter is to present the evaluation of the results. The accuracy assessment for the pixel-based classification and patch-oriented image categorization for considered feature descriptors will be carried out.

In the first section we present the overview of the framework in which the evaluation will be carried out. In the second section, we will briefly describe the evaluation criterion adopted, that includes details of the classification methodology and the evaluation measures.

Third section deals with the pixel-based classification accuracy assessment. The database for pixel-based classification evaluation is presented which consists of 12 homogeneous well textured metric-resolution SAR sub-images from TerraSAR-X. Later pixel-based classification results with the analyzing window of 16×16, 32×32 and 64×64 pixels for considered feature descriptors in context with the methodological framework from the first section will be presented and discussed.

Fourth section deals with the patch-oriented image categorization assessment. The overview of the database used is presented, which constitutes 2000 image-patches also from the metric-resolution TerraSAR-X images. The presented experimental database comprises of 20 well-defined objects / texture categories with 100 patches of each category. The size of each image patch is 200×200 pixels, thus each patch covers approximately $200^2 m^2$ on ground. The qualitative analysis of the feature descriptors is carried out by visualizing some of the planes from the feature space for each method followed by the quantitative assessment.

In the last part of the section, we propose the use of logarithmic-cumulants for further enhancing the accuracy of the FrFT-based and Gabor-based methods. Improvement in the category separability by using the logarithmic-cumulants is evident in the new feature space diagram. Quantitative accuracy assessment is discussed in detail for the FrFT-based and Gabor-based methods.

6.1 Overview: discussion framework

In section 3.1.1.2 we presented examples of diverse categories of textures and objects as image content in high-resolution SAR images. We also stated that depending upon the image content, spatial content understanding and in turn the application area will be defined. Well defined texture will be suitable to carry out pixel-based classification to achieve global classes, in the range of a total of $5-10$ global classes. Whereas, a complicated arrangement of objects in high-resolution with larger objects will lead to a more suitable area of application, i.e. patch-oriented image categorization. In each case, what is required is a very compact intrinsic multidimensional **feature descriptor** which describes the spatial content of the image. This **feature descriptor** can be obtained using parametric or nonparametric approaches, in image domain or an image within a transformation domain.

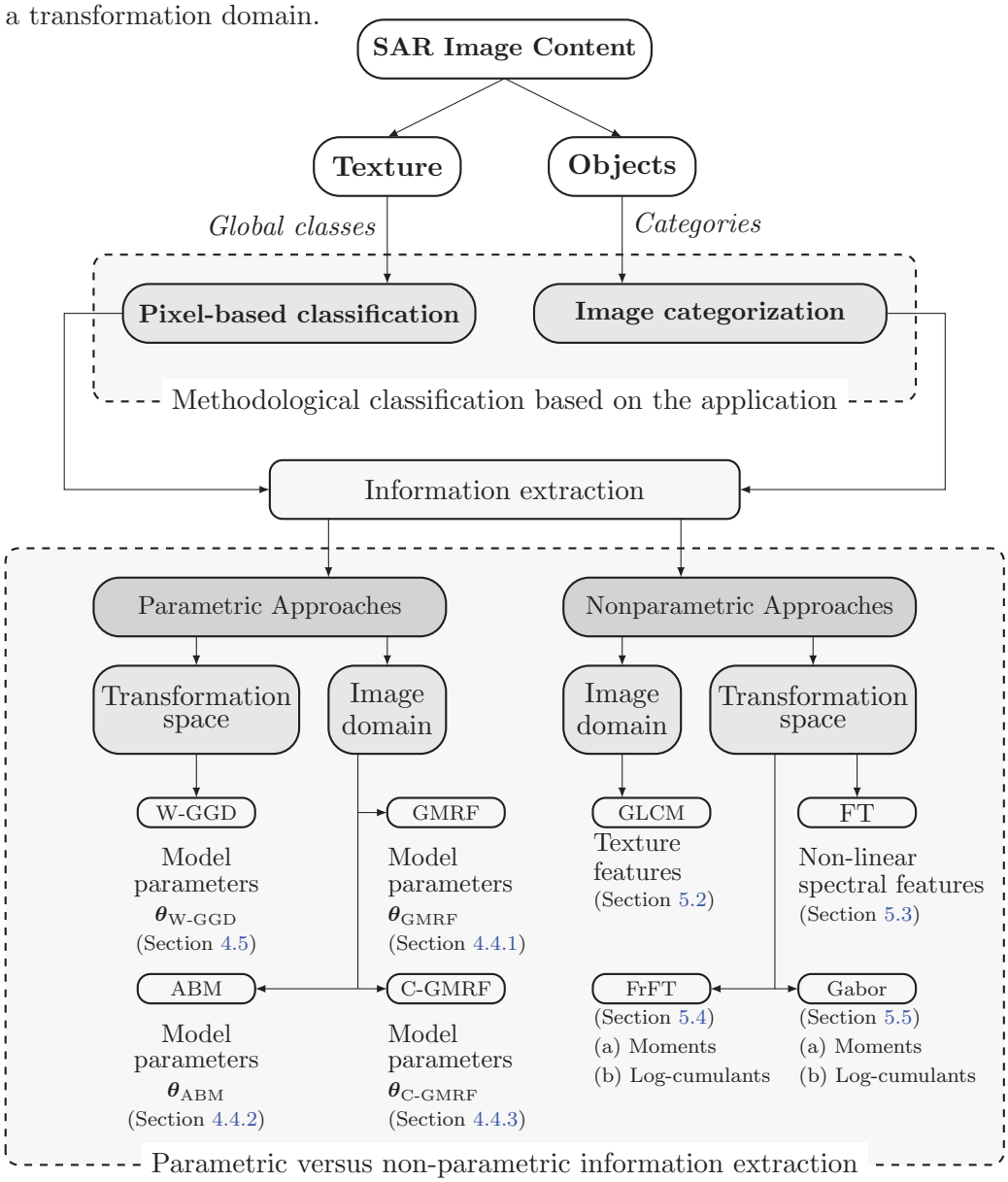


Figure 6.1: Qualitative performance of several feature descriptors from parametric and nonparametric class of methods will be assessed in an application framework based on the image content.

In this chapter, we will provide this parametric versus nonparametric feature descriptor analysis for the pixel-based classification and patch-oriented image categorization. We made a first attempt to carry out a comparative qualitative study of a parametric and a nonparametric approach for pixel-based classification in very-high-resolution SAR images in Singh *et al.* [2009]. We will enrich that discussion with more approaches and also encapsulating the patch-oriented image categorization issue. The considered parametric and nonparametric feature descriptors, as shown in the figure 6.1, have been discussed in detail in chapters 4 and 5, respectively. Following acronyms will be used hereafter in the graphical presentations of results for a better readability of results.

Parametric approaches

GMRF : Gauss-Markov random fields model as *prior* pdf for amplitude images.

ABM : Autobinomial model as *prior* pdf for amplitude images.

C-GMRF : Gauss-Markov random fields model as *prior* pdf for complex-valued images.

W-GGD : Generalized Gaussian distribution parameters for wavelet coefficients.

Nonparametric approaches

GLCM : Gray level co-occurrence matrix based statistical features for amplitude images.

NL-SFD : Non-linear spectral features for amplitude images.

FrFT : The fractional Fourier transform based features for complex-valued images.

Gabor : Gabor expansion based features for amplitude images.

6.2 Evaluation criterion

In this section, we present a general overview of the employed evaluation methodology (section 6.2.1) and evaluation measures (section 6.2.2) for the quantitative analysis of presented methodologies. A set of experimental database has been generated to quantify the pixel-based classification and patch-oriented image categorization performance.

6.2.1 Classification methodology

The clustering / classification can be performed by unsupervised and supervised classifiers. In the unsupervised clustering case, a conventional pattern recognition algorithm, e.g. k-means (MacQueen [1967]) or modified k-means, with a pre-defined number of clusters (k) can be used. In the case of supervised classification, the nearest neighborhood based pattern classification technique, e.g. k-nearest neighborhood (k-NN), where the regions of interest corresponding to the evaluated class are given can be employed (Cover & Hart [1967]).

Although the unsupervised clustering such as k-means are computationally less cumbersome, but there is no control over the specific identity of the classes, thus in order to provide a quantitative evaluation of various methods, we restrict ourselves to the supervised classification, e.g. k-NN classification. The k-NN method, first mentioned in Fix & Jr [1951], is among the simplest, however slow, instance-based learning method and very well accepted as a standard tool of classification. In this method, k is the user defined distance metric, based on which probability of class belonging to a class of labeled patterns is defined. Usually this distance metric is the Euclidean distance or the Mahalanobis distance (Mahalanobis [1936]), however some other metrics such as sum of absolute differences, or Cosine similarity measure may be

employed. Given the set of labeled pattern (training samples), the unlabeled samples in dataset are assigned the label which belongs to same category as its closest neighbors (Mitchell [1997]). The choice of the values of k, i.e. the number of neighbors taken into consideration while assigning a class label to an unlabeled sample is very empirical and depends upon the data. The value of $k = 1$ results into nearest neighborhood algorithm, which is also what we used in our experiments.

6.2.2 Confusion matrix and accuracy measures

The confusion matrix or the error matrix is frequently employed to organize and display information used to assess the thematic accuracy of a results of classification. A numerous accuracy measures have been proposed for summarizing the information contained in this confusion matrix (Stehman [1997]).

		Predicted classes					
	Class$_{i,j}$	$i = 1$	2	3	...	n	**Row Total**
Reference classes	$j = 1$	$w_{1,1}$			...	$w_{n,1}$	$\sum_{i=1}^{n} w_{i,1}$
	2		$w_{2,2}$		...	$w_{n,2}$	$\sum_{i=1}^{n} w_{i,2}$
	3			$w_{3,3}$	...	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	n	$w_{1,n}$	$w_{2,n}$	$w_{3,n}$	...	$w_{n,n}$	$\sum_{i=1}^{n} w_{i,n}$
	Column Total	$\sum_{j=1}^{n} w_{1,j}$	$\sum_{j=1}^{n} w_{2,j}$	$\sum_{j=1}^{n} w_{3,j}$	...	$\sum_{j=1}^{n} w_{n,j}$	$\sum_{c=1}^{n} w_{c,c}$

Figure 6.2: An example of a confusion matrix along with the *row total* and *column total* used to calculate the user's accuracy and producer's accuracy

The general formulation of a confusion matrix is shown in the figure 6.2. For a data with n number of classes, the confusion matrix will have the dimension of $n \times n$. The row indexes j of a confusion matrix correspond to actual or reference classes and the column indexes i correspond to predicted classes produced by applying the model. Each element $w_{(i,j,|i\neq j)}$ in the confusion matrix represents the number of samples from class i that were misclassified as belonging to class j, whereas the diagonal elements $w_{(i,j,|i=j)}$ in the confusion matrix represent the correctly classified samples. The sum of diagonal elements of the confusion matrix, $\sum_{c=1}^{k} w_{c,c}$, provides the total accuracy.

The confusion matrix provides a complete overview of accuracy of classification (both for individual classes and overall accuracy). When considering a quantitative comparison of many methods, the complete confusion matrix is not a very convenient tool to summarize the assessment at a glance, but still confusion matrix can be used as a basic tool to compute more descriptive accuracy measures. The total accuracy, which is the sum of the major diagonal elements of the confusion matrix may provide a quick summary, but it fails to reveal if the error was evenly distributed and it fails to quantify the performance of the feature descriptor for the individual classes. Thus, other forms have been derived from confusion matrix, i.e. the user's accuracy (UA) and producer's accuracy (PA) which are very descriptive and sufficient for a analytical overview (Story & Congalton [1986]).

The user's accuracy for a particular texture class (or image category) $\mathbf{c}$ is given by the ratio of the individual class (or category) accuracy $w_{c,c}$ and the total of the $\mathbf{c}^{th}$ row. The user's accuracy includes the error of commission and informs how well the classified results relates to the reference data. The producer's accuracy for a particular texture class (or image category) $\mathbf{c}$ is given by the ratio of the individual class (or category) accuracy $w_{c,c}$ and the total of the $\mathbf{c}^{th}$ column. The producer's accuracy includes the error of omission and informs how well a certain class / category has been classified. Usually both the accuracy measures are presented in the percentage and formulated for a particular class/category $\mathbf{c}$ as follows:

$$\mathrm{UA} = 100 \times \frac{w_{c,c}}{\sum_{i=1}^{k} w_{i,c}}, \quad c = 1, 2, \ldots, k. \tag{6.1}$$

$$\mathrm{PA} = 100 \times \frac{w_{c,c}}{\sum_{j=1}^{k} w_{c,j}}, \quad c = 1, 2, \ldots, k. \tag{6.2}$$

It must be noted that all the accuracy measures are only the estimates of the actual accuracy and are very much sensitive to the training samples used in generating the classification map. The size of training sample also plays an important role in the accuracy estimates. Thus it is always advocated to represent the accuracy assessment with variable sizes of the training samples.

We will consider the quantitative analysis of considered feature descriptors for pixel-based classification first. Later, qualitative as well as quantitative analysis of considered feature descriptors for patch-oriented image categorization will be presented. As the focus is shifting on patch-oriented image categorization increasing with the advent of high-resolution satellites, we will suggest improvement to the standard feature descriptors for enhancement of patch-oriented image categorization accuracies, specifically for Gabor filter banks and the fractional Fourier transform based methods.

6.3 Pixel-based classification

6.3.1 Database for pixel-based classification evaluation

For the pixel-based classification, we selected 12 TerraSAR-X sub-scenes of metric-resolution depicted in the figure 6.3. The size of each image is 256×256 pixels, i.e. approximately 256^2 m^2 area covered on the ground. Each image in the database is complex-valued, i.e. the mosaic can be used to test the algorithms specifically applicable to the complex-valued images. Each sub-scene covers a specific texture from man-made to natural land-cover typologies typically possible in the metric-resolution SAR images. A great care has been taken to have a homogeneous area-cover in each sub-scene, which is essential for the evaluation of pixel-based classification algorithms. The various textures can be identified as $\mathrm{T01}, \mathrm{T02}, \ldots, \mathrm{T12}$. Sub-scenes T01, T04, and T08 cover different forest areas, sub-scenes T02, T05 and T12 are differently textured open-fields / grassland area. Four sub-scenes, T03, T06, T07 and T10 cover various regions with man-made objects forming a texture, specific to the concerned locality / habitat. Sub-scene T09 is a pure water class from the calm-sea region, and T11 is the rough water class close the a coastal region. As the number of texture classes used is 12, the dimension of the confusion matrix for pixel-based classification shown in figure 6.2 will be 12×12. The lengths of the user's accuracy (equation 6.1) and producer's accuracy (equation 6.2) vectors will also be 12 each. The pixel-based classification is carried out usually with continuous displacement of analyzing window based operation, by computing feature descriptor on this analyzing window (aw). This small aw

is assumed to have a homogeneous texture. There is no standard approach to optimally select *aw* size, thus the analysis has been carried out on a variable *aw* size to measure the sensitivity of the algorithms to the size of *aw*. The used *aw* sizes are 16×16, 32×32, and 64×64 pixels, with a displacement step of 1 pixel in each case. The training size used for supervised classification is 01% and 10% of the total unlabeled samples.

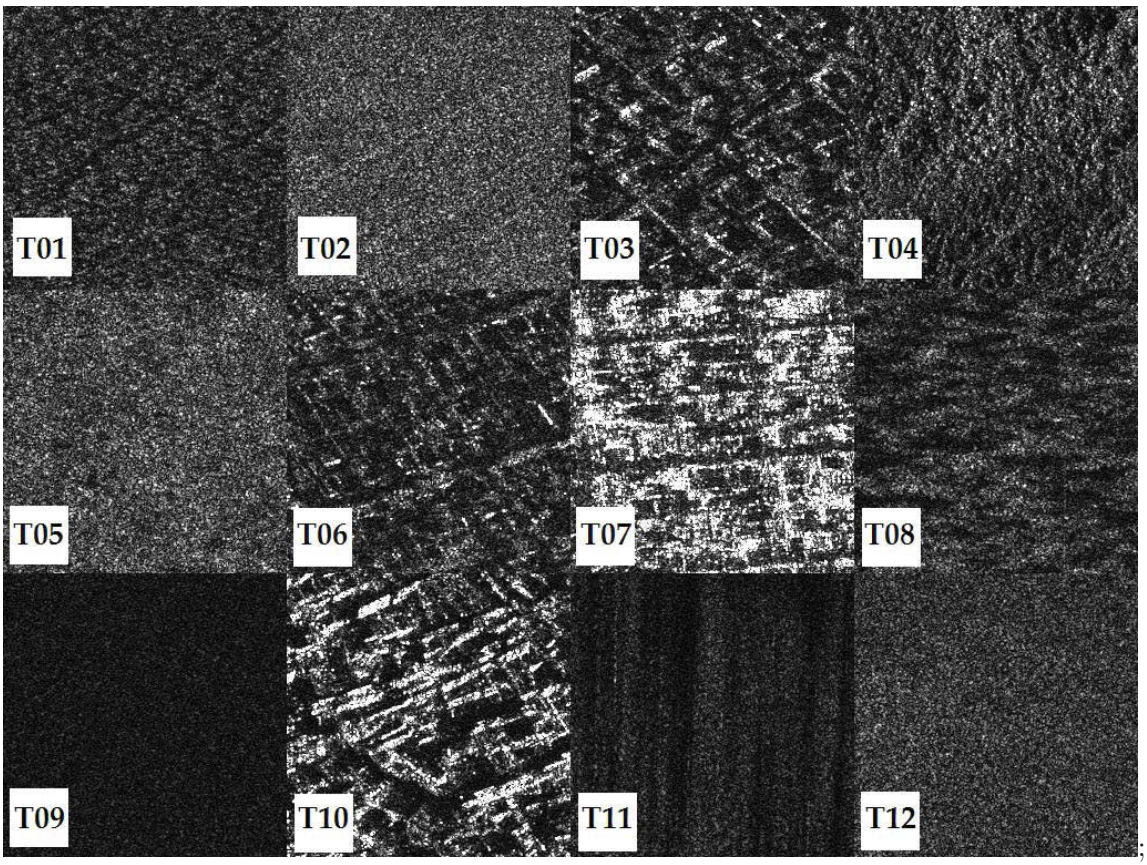


Figure 6.3: Amplitude representation of the mosaic of 12 metric-resolution SAR image subscenes covering various homogeneous textures. This includes textures in urban localities to the land-cover typologies as well as calm and rough water classes. The database generated is the complex-valued database, however amplitude images can be generated as specific to the algorithm requirement.

6.3.2 Results and discussions

The k-NN supervised classification have been carried out on the database of 12 TerraSAR-X subscenes of metric-resolution depicted in the figure 6.3. The size of *aw* has been chosen variably with 16×16, 32×32, and 64×64 pixels. The training samples for supervised classification are chosen to be 01% and 10%. So in each iteration, as a result of k-NN classification, a confusion matrix of size 12×12 will be generated. As the analysis of confusion matrix is very complicated, we will only present the user's and producer's accuracy. The average, and individual user's and producer's accuracies at 01% and 10% training sample case has been shown in the tables in the figures 6.4 and 6.5, respectively.

For these results we will not go into the details of individual texture accuracy assessment,

FD:		AA	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
GMRF	PA	**37.18**	33.89	30.98	34.73	52.09	28.45	33.10	50.62	26.29	40.17	54.54	30.33	30.97
	UA	**37.81**	33.47	27.94	37.32	51.06	26.43	34.42	57.83	28.16	37.92	61.75	29.22	28.16
ABM	PA	**39.44**	40.95	36.45	25.40	41.14	33.87	27.50	35.36	29.42	74.37	57.57	28.76	42.52
	UA	**40.18**	38.29	35.54	27.04	39.39	31.38	26.88	33.74	28.52	77.68	72.22	34.61	36.87
C-GMRF	PA	**16.13**	12.82	13.54	11.31	18.08	13.42	12.28	17.37	12.34	36.84	18.19	13.48	13.92
	UA	**16.35**	12.01	12.73	12.09	17.42	12.82	12.47	20.05	11.99	37.45	21.49	12.63	13.00
W-GGD	PA	**29.21**	22.60	41.23	16.66	16.39	30.32	16.26	29.68	15.28	68.43	29.18	34.45	30.07
	UA	**29.75**	19.42	39.37	16.17	15.72	28.41	15.75	34.86	15.21	71.07	30.34	39.20	31.47
GLCM	PA	**16.85**	12.95	17.10	14.65	11.74	15.35	15.09	29.47	10.48	18.88	28.97	11.65	15.87
	UA	**16.91**	12.79	16.23	14.72	11.58	14.39	14.99	29.56	11.01	19.80	30.49	12.20	15.20
NL-SFD	PA	**49.41**	37.87	55.56	24.26	45.20	41.83	23.36	56.52	29.59	90.23	43.38	79.45	65.64
	UA	**49.28**	35.63	53.49	24.39	46.19	41.47	24.12	59.02	31.82	90.41	42.40	78.58	63.86
FrFT	PA	**50.10**	38.21	53.36	29.13	40.33	43.34	26.33	61.07	36.30	88.27	49.16	75.45	70.27
	UA	**50.26**	36.77	54.83	30.81	38.77	42.57	25.84	64.15	25.86	88.02	48.97	75.38	71.13
Gabor	PA	**49.09**	34.35	49.01	30.91	37.74	34.09	29.65	67.90	33.08	91.89	52.65	74.40	53.41
	UA	**49.09**	33.01	44.20	33.64	39.79	32.66	30.21	71.14	33.79	86.86	54.89	78.44	50.42

(a) Analyzing window size: 16×16

FD:		AA	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
GMRF	PA	**67.19**	70.90	49.95	71.83	92.21	50.29	67.59	87.89	58.03	69.08	87.33	52.33	48.86
	UA	**65.04**	62.58	48.68	70.04	85.48	45.53	67.55	85.65	58.50	69.77	89.16	50.65	46.94
ABM	PA	**53.88**	51.84	45.26	49.01	62.33	33.38	41.42	46.23	56.39	77.99	67.18	79.79	35.73
	UA	**53.77**	53.26	46.51	44.49	67.27	30.85	42.91	50.17	56.07	63.81	66.56	79.59	43.74
C-GMRF	PA	**40.72**	31.52	34.43	31.48	52.53	32.17	32.20	40.18	31.40	93.03	37.12	36.97	35.64
	UA	**41.34**	27.96	30.04	36.22	50.61	29.56	35.96	52.11	31.41	90.16	47.03	35.00	30.01
W-GGD	PA	**50.82**	41.35	55.12	23.52	57.74	43.75	24.47	70.73	33.96	86.37	49.79	78.36	44.63
	UA	**51.29**	37.13	51.95	24.20	56.94	42.42	23.75	74.20	31.52	86.34	55.32	80.86	50.87
GLCM	PA	**28.19**	24.87	30.30	24.56	19.11	24.76	29.57	44.14	17.50	37.27	41.38	20.10	24.72
	UA	**28.43**	23.55	28.40	24.66	18.39	24.17	28.07	42.85	18.17	41.52	46.06	21.70	23.60
NL-SFD	PA	**70.84**	67.29	62.70	48.66	68.92	60.13	47.44	76.32	64.13	99.28	62.69	98.22	94.28
	UA	**70.89**	64.97	62.10	47.27	69.54	60.15	47.03	76.22	66.88	98.50	65.46	99.12	93.41
FrFT	PA	**74.59**	74.04	67.89	60.55	72.92	62.11	52.13	78.24	61.22	99.42	70.78	98.11	97.70
	UA	**74.59**	71.18	66.08	62.04	71.71	64.22	51.60	80.68	62.07	98.64	70.57	98.90	97.40
Gabor	PA	**77.64**	70.84	73.98	69.34	74.88	56.51	65.80	89.28	64.32	99.62	80.42	97.87	88.76
	UA	**77.72**	64.57	70.88	73.51	76.30	55.67	66.76	90.26	69.57	98.28	84.58	99.08	83.12

(b) Analyzing window size: 32×32

FD:		AA	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
GMRF	PA	**88.67**	93.18	81.01	91.61	99.68	78.71	91.28	98.39	89.27	92.61	96.56	79.93	71.85
	UA	**88.76**	90.69	76.85	92.78	99.57	76.65	91.58	97.52	91.13	95.29	98.38	82.37	72.32
ABM	PA	**53.88**	51.84	45.26	49.01	62.33	33.38	41.42	46.23	56.39	77.99	67.18	79.79	35.73
	UA	**68.68**	64.58	58.78	61.80	77.74	42.66	65.35	72.14	66.66	87.69	84.49	81.91	60.37
C-GMRF	PA	**75.60**	71.31	75.36	62.39	92.85	72.17	65.70	68.65	70.75	100.0	67.55	81.71	78.75
	UA	**75.96**	63.99	69.05	72.88	83.90	69.18	74.56	83.65	70.22	99.76	81.19	74.80	68.35
W-GGD	PA	**75.33**	64.10	64.65	58.23	88.88	60.53	44.89	91.65	66.27	95.26	89.40	92.92	87.20
	UA	**75.62**	55.38	64.59	60.50	88.61	59.62	48.42	92.80	63.32	96.23	91.35	96.50	90.07
GLCM	PA	**55.63**	57.80	63.12	52.10	50.36	51.94	58.32	71.01	38.68	70.35	65.74	40.67	47.50
	UA	**55.78**	55.34	56.43	55.11	50.71	48.03	52.43	66.23	40.78	77.34	72.94	45.61	48.39
NL-SFD	PA	**87.93**	89.97	77.32	71.91	87.78	74.26	75.29	96.22	86.84	99.99	95.75	99.90	99.98
	UA	**87.93**	86.74	75.03	73.48	86.26	76.60	74.82	96.38	89.88	99.90	96.15	99.99	99.99
FrFT	PA	**92.85**	96.10	88.87	82.87	95.62	86.29	83.59	94.91	89.87	100.0	96.09	100.0	100.0
	UA	**92.87**	92.59	86.64	86.02	94.06	88.58	81.17	96.41	94.38	100.0	94.59	100.0	100.0
Gabor	PA	**96.60**	97.87	93.14	95.07	95.77	89.43	94.78	98.03	97.39	100.0	98.27	100.0	99.50
	UA	**96.62**	97.74	92.11	96.68	96.88	87.69	95.14	98.66	97.44	100.0	97.63	100.0	99.50

(c) Analyzing window size: 64×64

Figure 6.4: Accuracy assessment of various feature descriptors for pixel-based classification. FD: Feature Descriptor, AA: Average accuracy, UA: Users' accuracy, PA: Producers' accuracy. Training sample size used: 01%.

FD:		AA	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
GMRF	PA	**64.81**	65.01	58.57	65.44	76.63	55.72	63.21	76.04	57.66	65.62	77.87	57.89	58.03
	UA	**65.21**	62.56	55.57	69.27	76.26	53.20	65.77	81.62	58.91	62.94	84.89	56.55	55.00
ABM	PA	**87.57**	87.70	86.62	84.43	87.75	86.34	85.16	87.22	85.44	95.27	91.78	85.17	87.94
	UA	**87.59**	86.44	86.94	85.42	87.53	84.50	85.69	86.17	84.86	95.55	94.18	87.88	85.96
C-GMRF	PA	**29.11**	24.81	25.27	25.11	30.17	25.47	25.78	33.60	24.53	49.39	33.15	25.99	26.01
	UA	**39.38**	23.88	24.22	26.57	29.90	24.50	26.52	36.73	24.41	48.50	36.94	24.77	24.37
W-GGD	PA	**38.85**	32.33	48.56	24.99	26.00	38.10	25.13	40.47	24.79	78.33	36.68	49.39	41.44
	UA	**39.55**	27.33	46.97	24.40	24.27	36.40	24.47	47.64	24.10	79.97	40.23	55.29	43.55
GLCM	PA	**24.80**	20.82	24.69	22.69	19.90	22.61	22.89	35.77	19.46	27.67	37.43	20.35	23.32
	UA	**24.83**	20.76	23.95	22.79	20.05	22.12	22.88	36.03	19.54	28.23	38.03	20.84	22.72
NL-SFD	PA	**55.21**	42.73	59.11	32.39	53.32	47.37	31.84	64.03	38.41	92.19	49.71	82.59	68.79
	UA	**55.10**	42.17	58.34	32.93	53.98	47.34	32.02	64.15	39.39	92.21	49.64	81.66	67.35
FrFT	PA	**57.92**	47.26	60.03	42.06	47.71	48.87	37.83	70.50	35.01	91.52	60.17	78.36	75.75
	UA	**57.91**	45.21	59.49	42.19	48.37	49.28	38.69	71.93	35.18	90.17	59.35	79.06	75.93
Gabor	PA	**61.17**	48.20	59.04	47.02	51.25	48.17	46.19	78.86	48.16	94.14	64.49	82.47	65.99
	UA	**61.04**	47.99	56.54	50.59	52.86	45.67	48.09	78.47	50.24	90.79	67.31	83.25	60.67

(a) Analyzing window size: 16×16

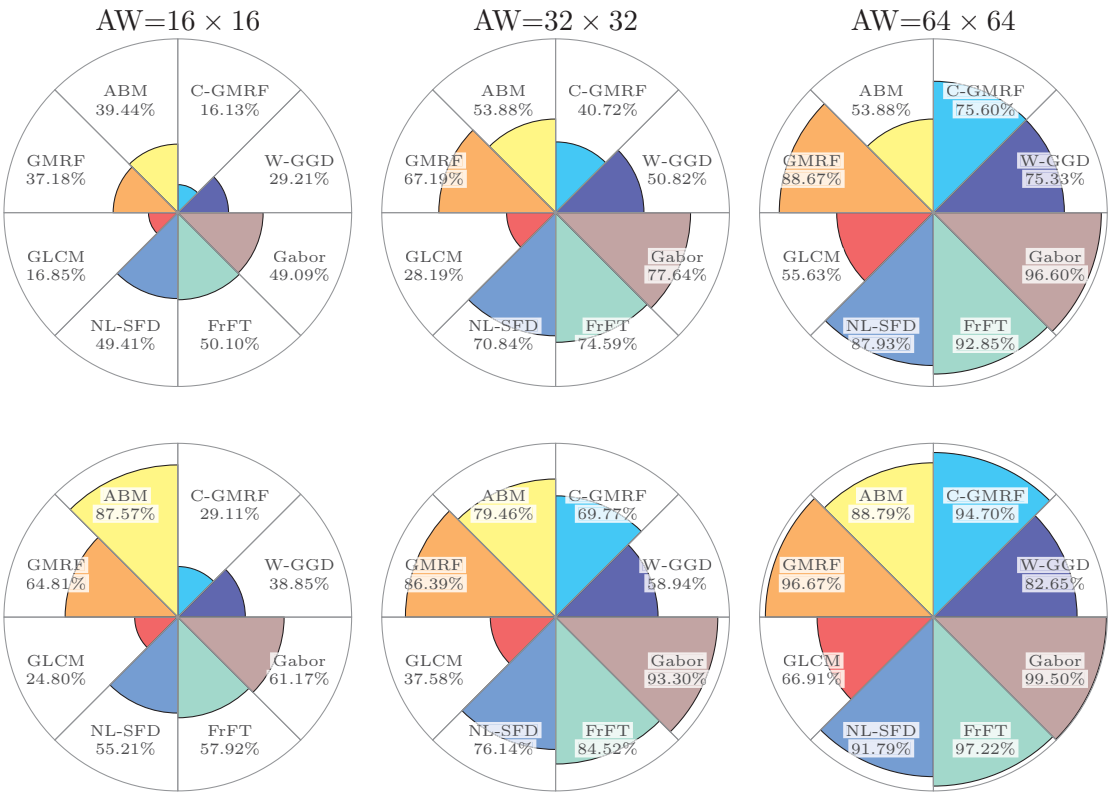
FD:		AA	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
GMRF	PA	**86.39**	90.26	77.34	89.06	97.61	76.72	88.08	97.66	83.52	87.77	96.64	76.16	75.81
	UA	**86.06**	85.42	76.34	90.50	95.93	74.64	89.09	96.57	84.54	89.09	97.46	77.67	75.47
ABM	PA	**79.46**	76.02	71.69	83.46	89.28	56.62	80.91	84.51	86.82	89.54	90.93	94.35	49.40
	UA	**79.38**	81.12	72.72	82.05	91.45	54.93	83.08	86.39	82.52	71.90	89.19	92.64	64.56
C-GMRF	PA	**69.77**	62.93	66.21	63.13	77.14	66.03	66.25	69.73	65.26	98.01	67.52	69.35	65.70
	UA	**70.22**	59.09	60.86	70.32	75.26	62.20	70.10	79.94	64.76	96.03	79.03	65.95	59.11
W-GGD	PA	**58.94**	49.36	61.83	33.63	66.99	51.69	34.15	76.34	41.78	90.04	60.71	83.99	56.71
	UA	**59.27**	43.72	60.82	34.91	65.80	51.57	34.08	79.44	39.63	90.57	64.09	85.99	60.64
GLCM	PA	**37.58**	32.81	37.39	34.80	28.85	33.53	38.11	51.91	28.19	47.20	53.32	31.49	33.30
	UA	**37.70**	32.27	36.81	34.54	28.31	32.79	37.37	51.51	28.56	49.13	55.94	32.68	32.44
NL-SFD	PA	**76.14**	72.00	66.46	57.63	75.59	65.61	58.76	81.11	71.35	99.46	70.90	98.92	95.86
	UA	**76.17**	71.09	66.94	57.38	75.72	65.12	57.46	81.50	72.32	99.08	72.80	99.32	95.26
FrFT	PA	**84.52**	82.31	74.61	78.53	84.16	73.06	73.26	90.19	73.64	99.76	86.75	99.21	98.73
	UA	**84.53**	79.63	74.56	78.96	83.50	73.44	71.92	91.51	76.28	99.46	86.92	99.54	98.62
Gabor	PA	**93.30**	91.78	91.97	90.46	93.02	85.75	89.19	97.64	87.54	99.95	95.22	99.70	97.38
	UA	**93.32**	88.00	89.21	93.36	92.87	85.84	90.35	97.78	91.81	99.78	96.43	99.90	94.51

(b) Analyzing window size: 32×32

FD:		AA	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
GMRF	PA	**96.67**	98.19	93.28	97.87	99.86	93.11	97.81	99.74	97.09	98.50	98.32	93.81	91.41
	UA	**96.68**	97.31	92.61	98.29	99.93	92.82	98.06	99.58	97.32	99.04	99.81	95.14	90.20
ABM	PA	**88.79**	84.38	83.45	92.66	94.10	61.45	91.82	92.07	92.27	97.33	98.66	97.69	79.58
	UA	**88.75**	90.27	78.66	90.71	96.18	66.51	90.56	94.87	91.97	95.17	97.23	97.08	75.83
C-GMRF	PA	**94.70**	95.40	96.14	89.17	99.14	95.69	91.70	89.43	95.11	100.0	91.07	96.89	96.71
	UA	**94.75**	91.78	93.03	94.77	95.83	92.96	95.10	96.67	93.87	99.96	95.89	94.64	92.47
W-GGD	PA	**82.65**	73.81	73.16	70.08	94.67	72.05	61.33	94.74	73.62	96.86	93.76	94.63	93.11
	UA	**82.85**	65.78	74.38	72.34	95.10	70.94	63.58	95.53	72.12	97.72	94.40	97.90	94.38
GLCM	PA	**66.91**	67.28	69.35	64.57	62.93	62.39	68.09	78.29	55.35	78.87	79.36	56.20	60.27
	UA	**66.97**	65.88	66.76	66.21	62.89	59.75	65.18	77.70	56.41	82.51	81.10	59.51	59.81
NL-SFD	PA	**91.79**	91.61	80.32	84.83	92.91	79.87	85.84	97.73	90.71	100.0	97.71	100.0	100.0
	UA	**91.80**	90.62	79.96	84.66	92.88	80.23	85.98	98.07	91.60	100.0	97.55	100.0	100.0
FrFT	PA	**97.22**	98.04	94.04	95.38	98.81	92.74	94.43	98.24	96.22	100.0	98.72	100.0	100.0
	UA	**97.22**	96.80	92.83	95.24	98.43	93.96	94.67	98.93	97.65	100.0	98.15	100.0	100.0
Gabor	PA	**99.50**	99.81	98.55	99.42	99.65	98.08	99.63	99.86	99.29	100.0	99.77	100.0	99.99
	UA	**99.50**	99.46	98.27	99.65	99.71	97.98	99.50	99.89	99.76	100.0	99.83	100.0	99.99

(c) Analyzing window size: 64×64

Figure 6.5: Accuracy assessment of various feature descriptors for pixel-based classification. FD: Feature Descriptor, AA: Average accuracy, UA: Users' accuracy, PA: Producers' accuracy. Training sample size used: 10%.



Producer's accuracies. First row- training samples: 01%, second row- training samples: 10%.

Users's accuracies. First row- training samples: 01%, second row- training samples: 10%.

Figure 6.6: Polar chart depicting the user's and producers's accuracies in the case of pixel-based classification for each feature descriptor at various analyzing window (AW) sizes. Left column- 16×16 pixels, center column- 32×32 pixels, and right column- 64×64 pixels.

but we will restrict ourselves only to the average accuracy assessment. The average accuracies (user's and producer's) are presented in form of polar charts for a clear understanding in the figure 6.6. The upper semi-circle in each polar chart depicts the parametric approaches, and lower semi-circle depicts the nonparametric approaches. We notice that there is no significant difference between user's accuracy and producer's accuracy in the results.

The user's accuracy includes the error of commission and is more complete representation of the confusion matrix, so lets assess only the user's accuracy. A quick glance at the polar charts shows that the accuracy improvement is achieved by increasing the number of training samples from 01% to 10%, but that is very obvious in well behaving features descriptors for supervised classification. Another important observation is that almost in every case, accuracy improvement is achieved when the analyzing window size is increased from 16×16 to 64×64, be it the parametric approaches or nonparametric approaches.

We first consider the case of nonparametric approaches, i.e. **GLCM**, **NL-SFD**, **FrFT**, and **Gabor**, depicted in the lower semi-circle in every polar chart. We see that **GLCM** based feature descriptor, a nonparametric approach in spatial-image domain, provides the poorest accuracy. The reason for this is the quantization of gray values in preprocessing step, as shown in the figure 5.3. We have quantized the images to 16 gray levels to achieve a computationally feasible analysis. Using fewer gray level is equivalent to looking at the image at a coarser level, which however seems to destroy the textural information in case of high-resolution images. But for a smaller window, it is important to restrict the number of gray-levels to obtain a statistically reliable estimate of the joint probability distribution in the co-occurrence matrices. This can be achieved by using a relatively large window, that's the reason why we obtain better accuracy at the *aw* size of 64×64 pixels than 16×16 pixels at same quantization level of gray levels. But increasing the window size even more will risk into uncertainties and errors as the texture changes over the large window, this issue will be evident in the patch-oriented image categorization in next section. It is not a practical alternative to increase the number of gray levels due to computational constraints. So we can conclude that its not advisable to use **GLCM** based method for pixel-based classification in high-resolution SAR images, and we will exclude **GLCM** in further discussions.

We observe that the accuracy enhancement is not considerable in case of nonparametric approaches (**NL-SFD**, **FrFT**, and **Gabor**) when increasing the training sample size from 01% to 10% for the case of 16×16 analyzing window size. All these methods work on the image within a transformation space. In the case of **NL-SFD**, for the *aw* of size 16×16 pixels, the sample size for the computation of the spectrum is too small for an accurate frequency estimate. Similar is the case for the **FrFT**, here also sample size at the *aw* size of 16×16 pixels is too small for an accurate joint time-frequency representation. In case of **Gabor**, going down in the stages of Gabor kernel, the sample size becomes even smaller and accurate bandpass filtering is not achieved. This observation is justified by the better accuracies in the *aw* of size 32×32 and 64×64 pixels. So with a sufficiently large window size, these nonparametric approaches applied on image within a transformation space are advisable for the texture analysis in the high-resolution SAR images for performing pixel-based classification.

Now we consider the case of parametric approaches, i.e. **GMRF**, **ABM**, **C-GMRF**, and **W-GGD**, depicted in the upper semi-circle in every polar chart. First three are based on texture model parameter estimation within spatial-image domain using Bayesian framework, selecting the model order 5. The fourth one, **W-GGD**, is based on the density parameter estimation for image within a transformation space, i.e. the Wavelet domain. For **W-GGD**, the number of levels used in the wavelet decomposition are 1, 2, and 3 for the *aw* of size 16×16, 32×32 and 64×64 pixels, respectively. Similar to earlier transformation domain methods in nonparametric

approaches, the performance of **W-GGD** also depends upon the sample size use. The *aw* of size 16×16 pixels seems to be very small to have enough number of samples for an accurate parameter estimation for assumed generalized Gaussian distribution for wavelet coefficients. This fact is again emphasized with better accuracies in the *aw* of size 32×32 and 64×64 pixels. But the accuracies at the *aw* of size 64×64 pixels for **W-GGD** is still significantly lower than the other well performing feature descriptors. The reason is again that the wavelet based multi-scale approach reduces the size of the image at successive levels, leading to risk of inaccurate density parameter estimation.

As far as the Bayesian inference based texture model parameter methods are concerned, the accuracies for each method are encouragingly good. It is certain that the higher the model order is, better the model can represent the spatial interaction between the pixels in an analyzing window leading to a better texture synthesis. But it must also be mentioned that the parameter computation is very computationally expensive task with the successive increase at the model order. Thus it might not be practically viable to go for a very high model order, so we restrict our analysis to model order 5. **ABM** provides best performance at the *aw* of size 16×16 pixels, followed by **GMRF** approach which are even better than the nonparametric approaches. Thus the Gibbs random fields based methods applied directly in the image domain are potentially very good methods for discriminating texture even at smaller analyzing window sizes. The requirement of a minimum number of samples for an accurate transformation in a suitable domain for nonparametric approaches applied on the images within a transformation domain is not existing in the Gibbs based methods, leading to a better accuracy even at smaller analyzing window sizes. The accuracy remains in the same range with increase in the *aw* size from 16×16 to 64×64 pixels at 10% training samples case.

Summary If we look at a overall scenario, following conclusions can be drawn with the demonstrations on experimental database used, in context to the methodological framework presented in the figure 6.1:

- Considering the case of the *aw* of size 64×64 pixels at 10% training samples, there is no significant difference in accuracy assessment in parametric and nonparametric approaches, excluding the case of **GLCM**. At a smaller *aw* of size 16×16 pixels, it is possible to achieve good accuracy in Gibbs random fields based parametric approaches employed at spatial-image domain, even at model order 5. Increasing the model order in parametric approaches will certainly improve the accuracy but at very high computational cost. A trade off must be made while selecting the model order for an appropriate method.

- At a smaller *aw* size of 16×16 pixels, the methods working on images within a transformation space are not very accurate due to a small sample size for accurate feature extraction.

- Methods based on image in transformation space perform better with a sufficiently large size of *aw*. But increasing the size of the *aw* will lead to heterogeneity in area of interest for texture synthesis and the case will be closer to an patch-oriented image categorization application.

- On the other hand, increasing the size of *aw* for Gibbs based parametric methods requires the use of higher model order, which is certainly not a viable choice considering exceptionally high computational cost for practical applications.

6.4 Patch-oriented image categorization

6.4.1 Database for patch-oriented image categorization evaluation

For the patch-oriented image categorization accuracy assessment, we generated a database of 2000 image-patches also from the metric-resolution TerraSAR-X images. The size of each image patch is 200×200 pixels, thus each patch covers approximately 200^2 m^2 of ground. This size can be generally used to define a particular category, as it is large enough to contain the contextual information needed to define an objects' structure in the case of very-high-resolution (Singh & Datcu [2012d], Popescu *et al.* [2012]).

The presented experimental database comprises of 20 well-defined objects / texture categories with 100 patches of each category. Care has been taken to generate the database with diverse textures. Examples of one patch from each category are shown in the figure 6.7. These categories include land-cover typologies (C01-*Grassland*, C02-*Forest*, C03-*Mixed vegetation*, C04-*Mixed vegetation with a water-channel* and C18-*Sandy plains*), urban area with a particular texture (C05, C06, C07 and C08-*Urban- type 01, 02, 03 and 04*), urban and natural water channel (C09 and C10), regions with strong scatterers along with the surfaces exhibiting specular reflection (C11-*Urban coastline* and C12-*Aircraft stands*), regions with very-strong scatterers with multiple bounces (C13-*Skyscrapers* and C14-*Industrial complex*), other special categories such as C15-*Airstrip*, C16-*Highway* and C17-*Train tracks*. Two categories from the sea-water region (C19-*Buoy in sea* and C20-*Ambiguities*) are also considered. As this database has been generated manually, so it can be used as a ground-truth for the assessment of various algorithms. Similar to the database for pixel-based classification, this database is also complex-valued, whereas on amplitude representation is shown here. In order to demonstrate the diversity of objects and texture in various categories, more examples from each category have been presented in the appendix in Fig. B.1.

The number of object categories used is 20, so the dimension of the confusion matrix for patch-oriented image categorization shown in the figure 6.2 will also be 20×20. The lengths of the user's accuracy (equation 6.1) and producer's accuracy (equation 6.2) vectors will also be 20 each. In patch-oriented image categorization, feature descriptor is computed for the whole SAR image patch. In other words, the size of analyzing window (aw) is same as the size of SAR image patch. The training size used for supervised classification is 1% and 10% of the total unlabeled samples.

6.4.2 Patch-oriented image categorization: results and discussion

Qualitative analysis: Before going into the quantitative analysis of results, we will perform a qualitative analysis by viewing some selected planes in the feature space for each feature descriptor. Figure 6.8 shows the feature space diagram for the parametric approaches and the figure 6.9 shows the feature space diagram for the nonparametric approaches. The feature space is essentially multi-dimensional, with dimensionality equal to the length of the feature descriptor. However, various feature descriptor will behave differently in other planes in the feature space, visual inspection of at least few planes in feature space diagram will provide an intuitive information about the performance of particular feature descriptor. In the figures 6.8 and 6.9, only a selected plane (i.e. only selected two dimensions) for each feature descriptor are shown to demonstrate the category-separability potential of each method. Normalized feature descriptors by limiting all of them to $[0, 1]$ for better visual inspection are presented.

It can be seen that, even if its a parametric or nonparametric approach, the feature descriptors generated on the SAR image within a transformation domain (**W-GGD**, **NL-SFD**, **FrFT**,

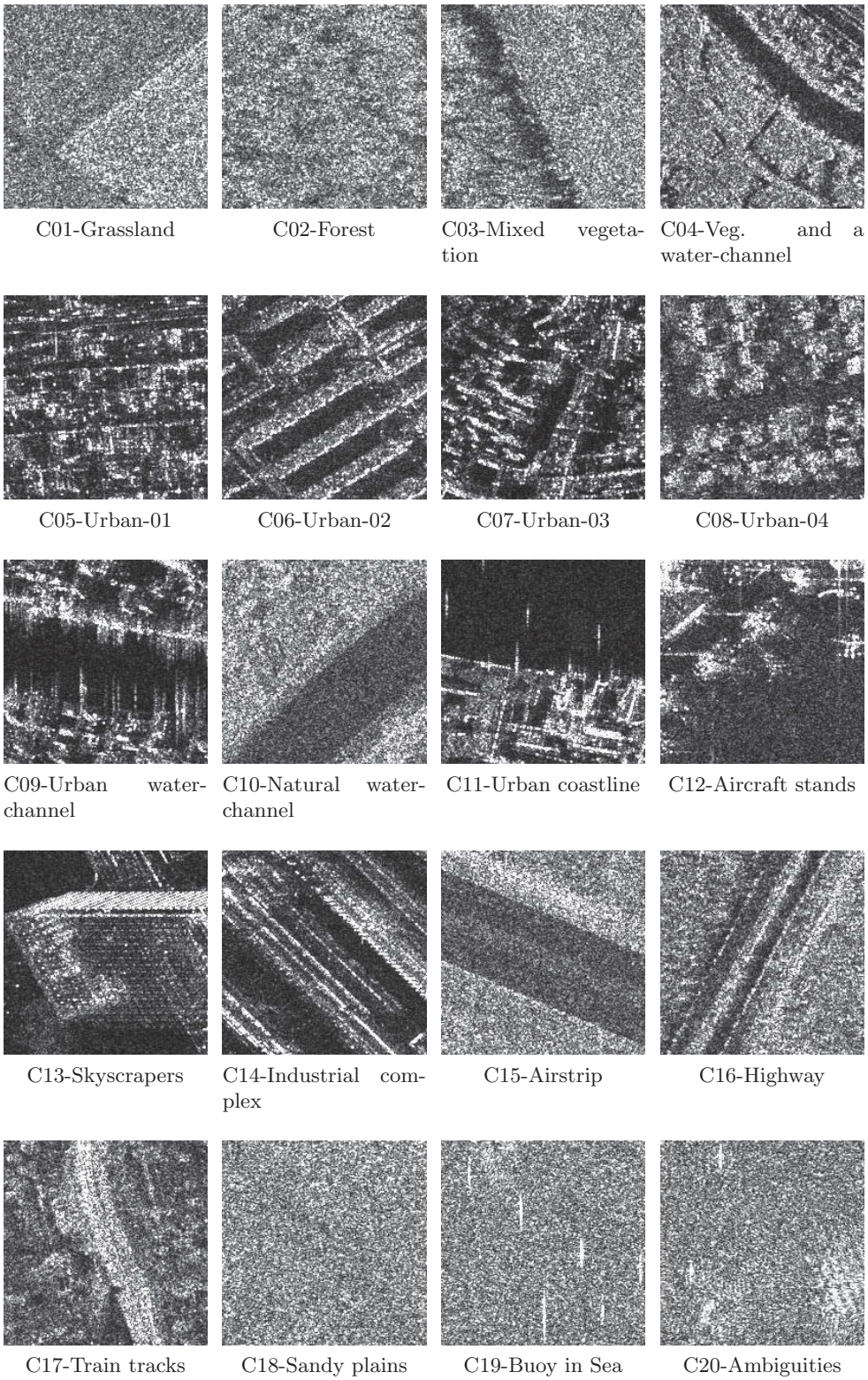
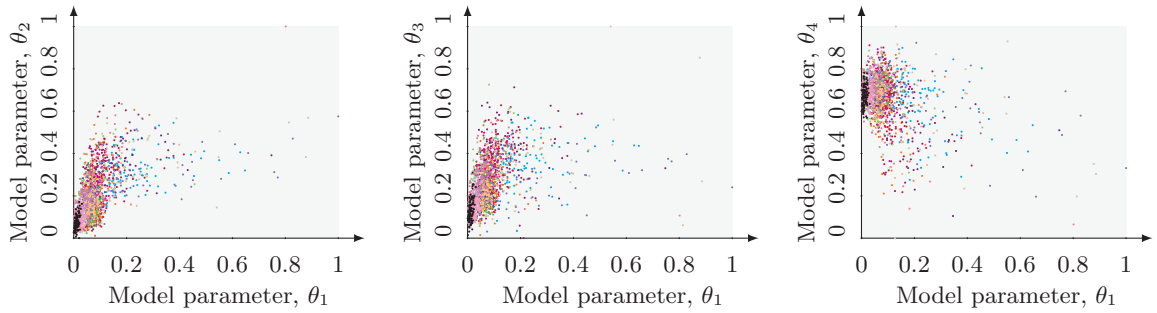
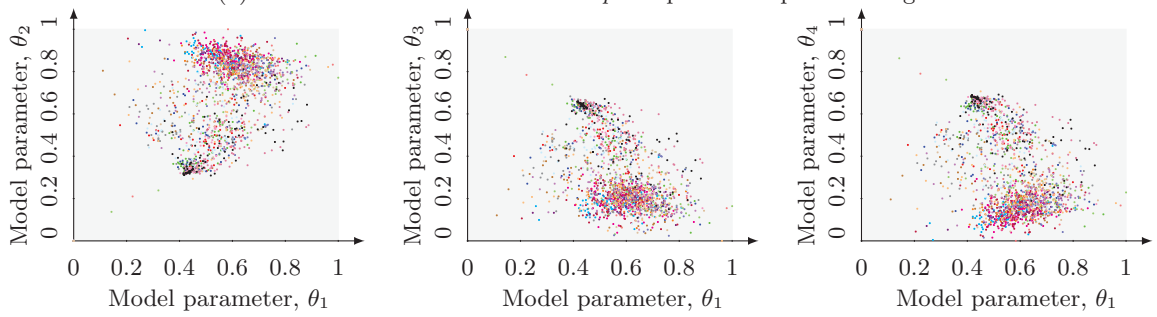


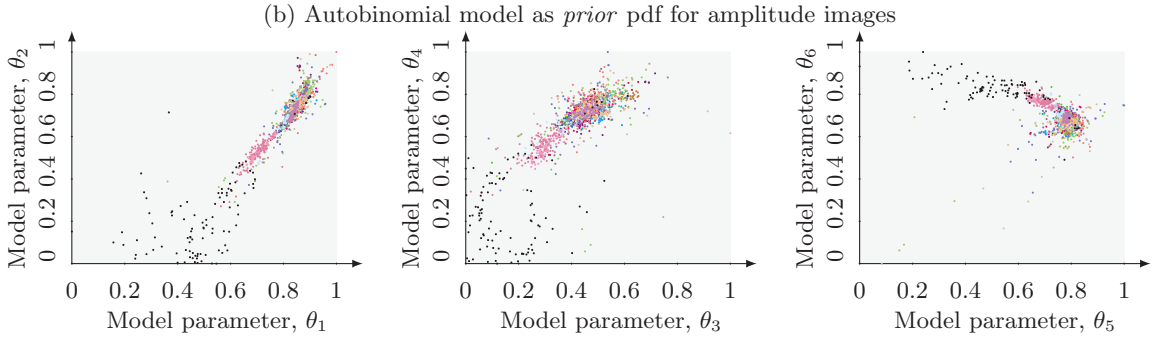
Figure 6.7: Examples of the SAR image patches (of the size 200×200 pixels) in different objects / texture categories. Image patches have been obtained from single-look complex very-high-resolution Spotlight TerraSAR-X images from different areas in order to include the diversity of objects / textures when analyzing feature descriptors.



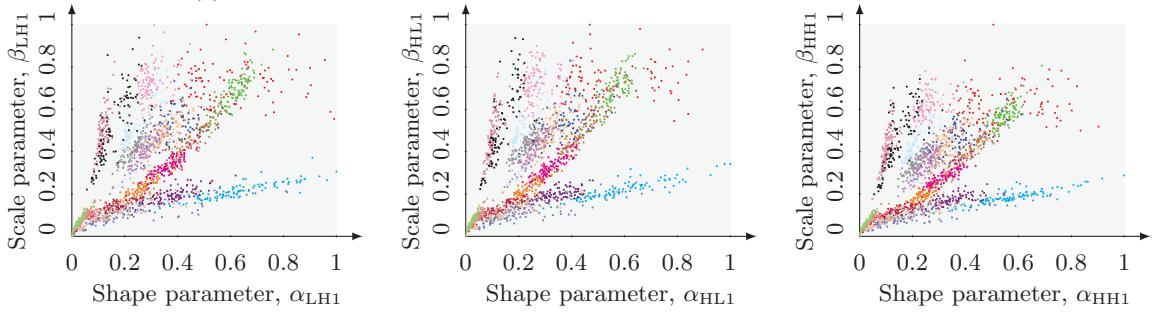
(a) Gauss-Markov random fields as *prior* pdf for amplitude images



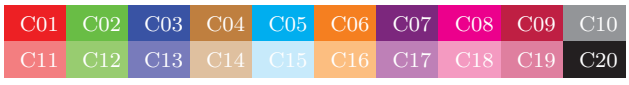
(b) Autobinomial model as *prior* pdf for amplitude images



(c) Gauss-Markov random fields as *prior* pdf for complex-valued images

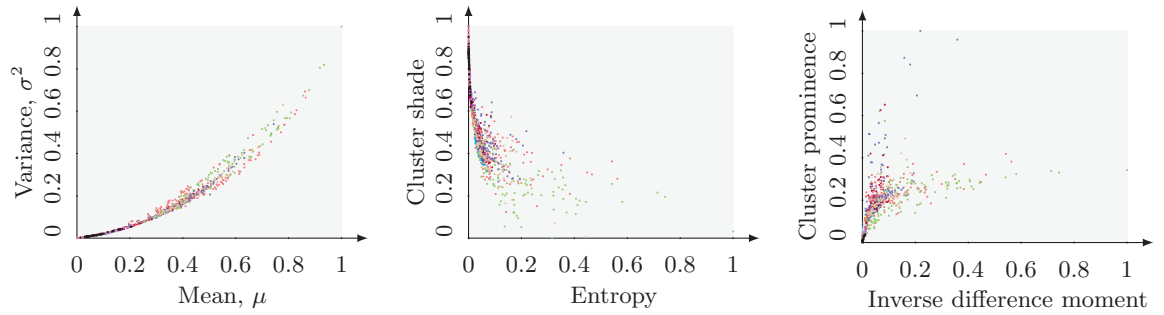


(d) Generalized Gaussian distribution parameters for the wavelet coefficients

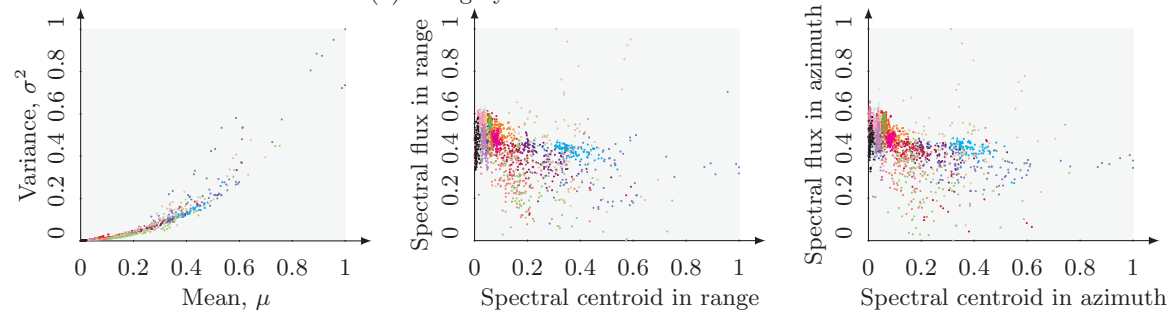


Category color-index

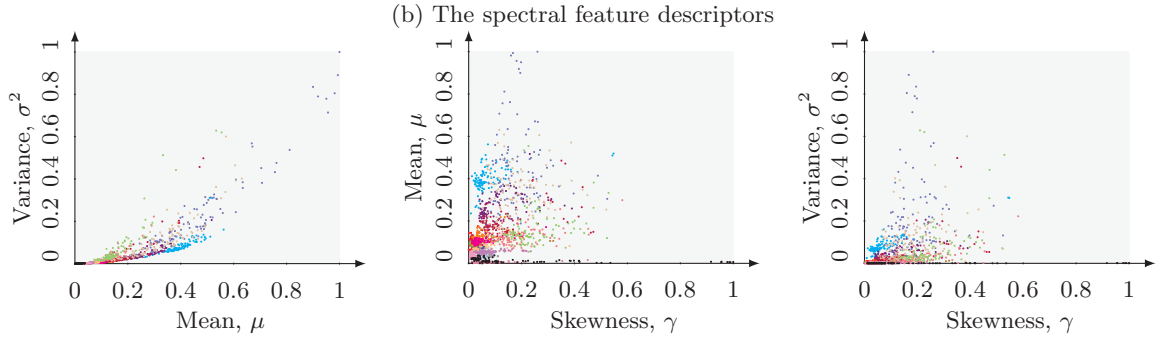
Figure 6.8: Feature space diagram demonstrating the potential of the considered parametric feature descriptors for category separability. Selected 2-D planes in the feature space for all feature descriptors are presented, however, it must be noted that the actual feature space is multi-dimensional. The dimensionality of feature space is same as the length of respective feature descriptor. Some features have very large positive or negative numerical values, thus for a better interpretation, all the features have been scaled between 0 and 1.



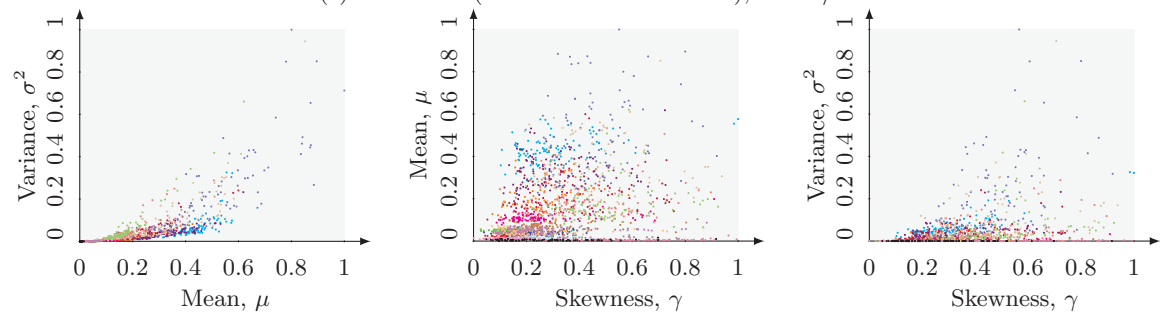
(a) The gray level co-occurrence matrix



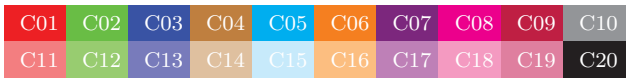
(b) The spectral feature descriptors



(c) The FrFT (The first kind statistics), $\alpha = \pi/8$



(d) Gabor (The first kind statistics), Scale: 1, Orientation: 2



Category color-index

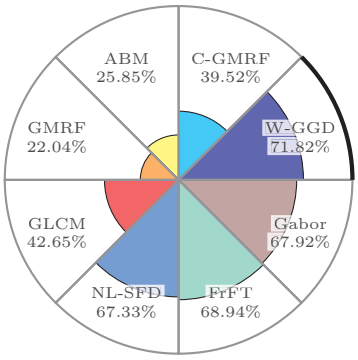
Figure 6.9: Feature space diagram demonstrating the potential of the considered nonparametric feature descriptors for category separability. Selected 2-D planes in the feature space for all feature descriptors are presented, however, it must be noted that the actual feature space is multi-dimensional. The dimensionality of feature space is same as the length of respective feature descriptor. Some features have very large positive or negative numerical values, thus for a better interpretation, all the features have been scaled between 0 and 1.

and **Gabor**) are showing better separability of categories in the feature space. Whereas, in case of the feature space of the feature descriptors generated directly on the SAR image (**GMRF**, **ABM**, **C-GMRF**, and **GLCM**), its very difficult to discriminate between various categories. This is first hand analysis to expect a better quantitative accuracy in case of **W-GGD**, **NL-SFD**, **FrFT**, and **Gabor** methods.

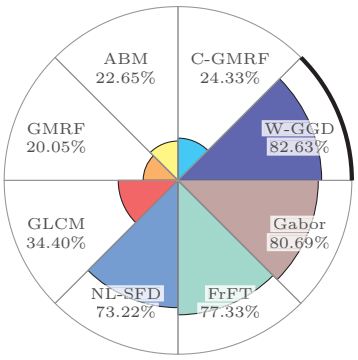
Quantitative analysis: Similar to the earlier case of pixel-based classification, we will only consider the user's accuracy for the quantitative analysis, as its a more complete representation of the confusion matrix. We will consider only the case of k-NN classification with 10% training samples. Apart from mean accuracy assessment, we will also carry out the individual category accuracy assessment. Polar charts depicting the mean accuracy (first polar chart) and individual category accuracies (further polar charts) are presented in the figure 6.10.

It can be noticed from the first polar chart in the figure 6.10 that the best mean accuracy is obtained in the **W-GGD**, followed by **FrFT**, **Gabor**, and **NL-SFD**. We also comment on the case of methods with considerably poor accuracies. The accuracies are very poor for the methodologies **GMRF**, **ABM**, **C-GMRF** and **GLCM**. These results are in coherence with the visibility of class separability in the various planes of the features spaces shown in the figures 6.8 and 6.9. The reason for poor accuracy in **GLCM** is again the same that the 16 quantization level to make the feature descriptor generation a practically viable task, essentially leads to a loss of most of the information. The other methods with poor accuracies, i.e. **GMRF**, **ABM** and **C-GMRF** are based on the texture parameter estimation with an assumption of a *prior* pdf. This assumed *prior* pdf is suitable for homogeneous regions, but fails to capture the heterogeneity in case of image-patches with a complex arrangement of objects. Moreover, these methods are based on an assumed multiplicative speckle noise model, with representing a SAR image as $y = x \times z$. Where y represents a SAR image, x denotes an ideal noise-free image, and z stands for speckle. Here we would also like to question the applicability of multiplicative speckle noise model, in case of image patches containing a large number of strong scatterers. So the situation is not such simple at the regions in SAR image patches with strong scatterers where the definition of speckle needs to be restated. Here speckle is not noise any more but the actual useful signal. So particularly for such areas, SAR image should be represented as $y = x + n$, where n is the noise (not speckle) which can be ignored for further considerations. For such scenarios we need to reconsider the utility of the Gibbs based methods. So we will ignore these 3 Gibbs based methods, i.e. **GMRF**, **ABM**, and **C-GMRF**, as well as the **GLCM** in further discussions. However, we would like to mention very good performance of **C-GMRF**, particularly for the categories 18, 19, and 20. The use of complex-valued data leads us to achieve better accuracies in these cases due to importance of the phase information in these category, which will be evident in further discussion on **FrFT**.

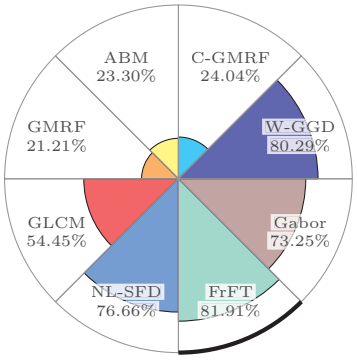
The difference between the accuracies of **W-GGD**, followed by **FrFT**, **Gabor**, and **NL-SFD** is very marginal. Thus we cannot arrive to a general agreement on an overall best analysis method that outperforms all the others on various tasks such as texture-based or backscattering behavior-based retrieval, segmentation, or classification, the notion earlier also discussed in literature Kandaswamy *et al.* [2005]. This leads to a need to carry out the individual category accuracies for each category for several feature descriptors. As mentioned in chapter 5, the feature descriptors for **FrFT** and **Gabor** constitutes the moments of the sub-bands achieved in each case, which is a very obvious choice. In the next section, along with the detailed analysis of individual category accuracies, we will propose an enhanced feature descriptor which constitutes the log-cumulants of the sub-bands to improve the overall accuracy over moments.



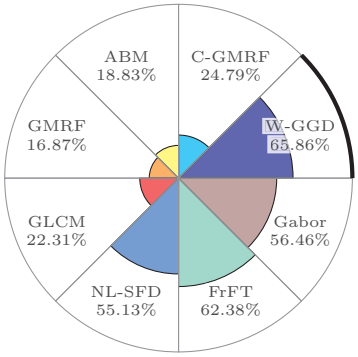
(a) Mean accuracy



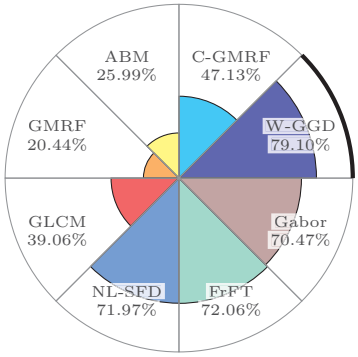
(b) Category 01



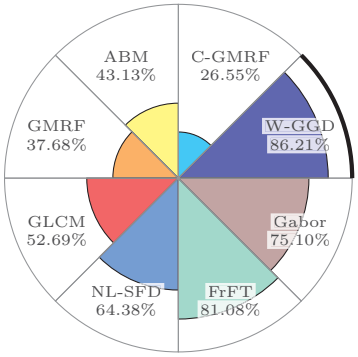
(c) Category 02



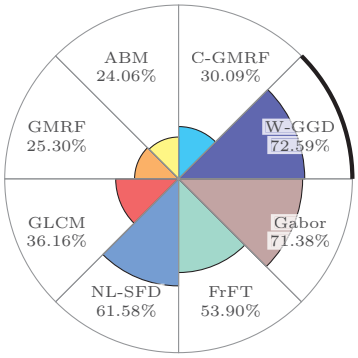
(d) Category 03



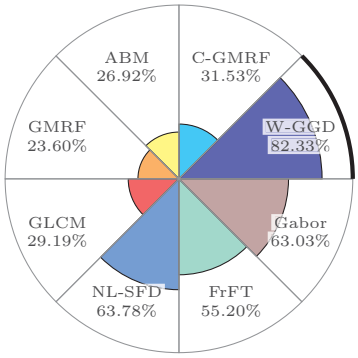
(e) Category 04



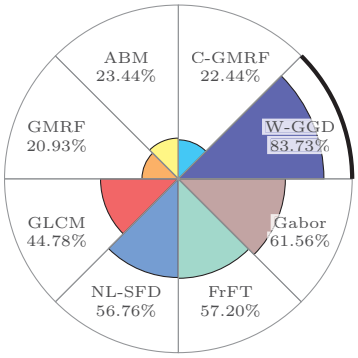
(f) Category 05



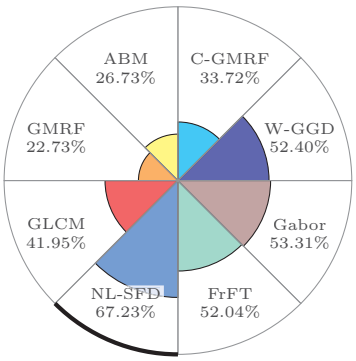
(g) Category 06



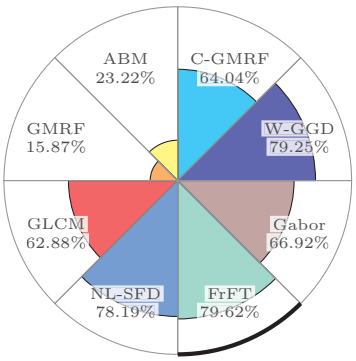
(h) Category 07



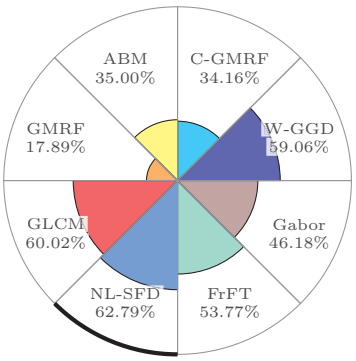
(i) Category 08



(j) Category 09

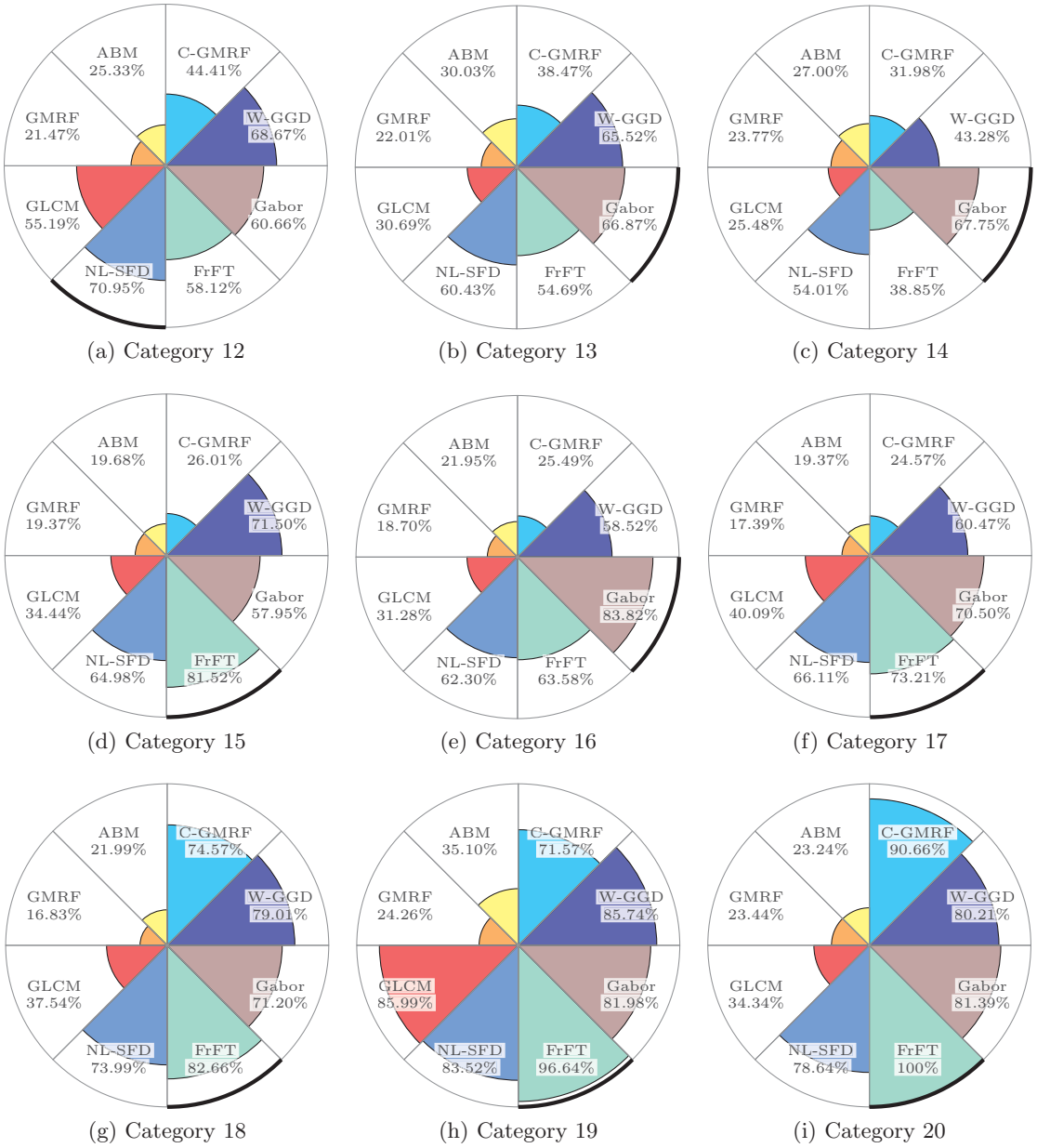


(k) Category 10



(l) Category 11

contd.



(a) Category 12 (b) Category 13 (c) Category 14

(d) Category 15 (e) Category 16 (f) Category 17

(g) Category 18 (h) Category 19 (i) Category 20

Figure 6.10: Polar charts depicting the individual category user's accuracies in the case of patch-oriented image categorization for each feature descriptor at 10% training samples. The feature descriptor with highest accuracy in each category has been highlighted with an bold arc on the outer circle representing 100% accuracy level. First polar chart on the previous page represents the overall mean accuracy. It is evident from the overall mean accuracy assessment that the methods employed directly in the image domain (parametric (**GMRF**,**ABM**, and **C-GMRF**) as well as nonparametric(**GLCM**)) provides significantly poor overall accuracy than the methods employed on the image within a transformation domain (parametric (**W-GGD**) as well as nonparametric(**NL-SFD**, **FrFT**, and **Gabor**)). Individual accuracy assessment suggests that no feature descriptor outperforms all other in each individual category scenario, thus makes it difficult to arrive at a general consensus on a best performing method.

6.4.3 Use of 'logarithmic-cumulants' for accuracy enhancement

We noticed that the methods pertaining to image within a transformation space (**NL-SFD**, **FrFT**, **Gabor**, and **W-GGD**) were providing better accuracies than the methods employed directly in the spatial-image domain (**GLCM**, **GMRF**, **ABM**, and **C-GMRF**). One important reason for this is that images represented in suitable transform space reveals certain latent information, otherwise difficult to exploit in spatial-image domain. Such as the FrFT method provides information about the backscattering behavior and Gabor filter banks ease texture synthesis. A feature descriptor have been generated by computing the moments of the amplitude of the FrFT coefficients as well as Gabor filtered images in each case respectively. In this section, we will suggest a methodology to further improve the overall accuracy particularly for these two cases, which has been motivated from the method of logarithmic-cumulants.

Second important reason for working in transformation space is that normally via the transformation of the image's pixel value into a suitable space, it is possible to obtain a more uniform statistical response that is easier to analyze (Do & Vetterli [2002]), such as wavelet coefficients in each sub-band can be approximated well with generalized Gaussian distribution. This makes the parameter estimation an easier task. Parameter estimation for the models of image pdf or transformed image pdf can be carried out using the classical maximum likelihood (ML) or the method of moments (MoM) approaches. The method of logarithmic-cumulants (MoLC) has also been proposed for efficient parameter estimation with pdfs defined on $\mathbb{R}^+$, which is also the case for the amplitude of the FrFT coefficients. The MoLC method is based on the analytical computation of logarithmic-moments and logarithmic-cumulants as a function of unknown parameters. The logarithmic-moments and logarithmic-cumulants have been introduced as 'second kind statistics' for SAR image analysis where the characteristic function is defined in terms of the Mellin transform instead of the Fourier transform. A detailed introduction to MoLC is presented in appendix C. The first three 'logarithmic-cumulants' are same as the first three respective 'logarithmic-moments'. Even though the MoLC provides an excellent method for parameter estimation, the problem associated with image patches in high and very-high-resolution SAR images persists even after the transformation (such as in case of **FrFT**, and **Gabor**), as it is hard to predict which pdf model fits the data in order to carry out parameter estimation.

Motivated from MoLC, we suggest to use the logarithmic-cumulants instead of moments to generate a feature descriptor for **FrFT** (Singh & Datcu [2012d]), and **Gabor** (Singh & Datcu [2012d]). Important to mention that we do not postulate any particular model for the distribution of the amplitude of the FrFT coefficients or Gabor filtered images, unlike the generalized Gaussian distribution modeling for the wavelet coefficients. We restrict ourselves only to the first stage in the MoLC, i.e. computation of sample logarithmic cumulants. This avoids the problem of selecting a pdf with sufficient goodness of fit. The first three logarithmic-cumulants (logarithmic-mean $\tilde{\kappa}_1$, logarithmic-variance, $\tilde{\kappa}_2$, and logarithmic-skewness, $\tilde{\kappa}_3$ estimates) for an image or image in transformation domain $I = \{r_1, r_2, \cdots, r_{\mathrm{N}}\}$ are given mathematically as (Li *et al.* [2010], Papoulis [1991]):

$$\tilde{\kappa}_1 = \frac{1}{\mathrm{N}} \sum_{k=1}^{\mathrm{N}} \ln r_k, \quad \tilde{\kappa}_2 = \frac{1}{\mathrm{N}-1} \sum_{k=1}^{\mathrm{N}} (\ln r_k - \tilde{\kappa}_1)^2, \quad \tilde{\kappa}_3 = \frac{\mathrm{N}}{(\mathrm{N}-1)(\mathrm{N}-2)} \sum_{k=1}^{\mathrm{N}} (\ln r_k - \tilde{\kappa}_1)^3. \quad (6.3)$$

With ln as the natural logarithm. The origin of these formulations with their background and importance is discussed in following sub-sections.

6.4.3.1 Why logarithmic-cumulants ?

Now the question arises, why the logarithmic-cumulants should be preferred over the moments ? If we denote $\tilde{m}_i$ as the i^{th} order logarithmic-moments for a random variable u with probability density function $p_u(u)$, following the traditional terminology from the moments and cumulants (Stuart & Ord [1987]), first three logarithmic cumulants can be written as follows:

$$\tilde{\kappa}_1 = \tilde{m}_1 \tag{6.4}$$

$$\tilde{\kappa}_2 = \tilde{m}_2 - \tilde{m}_1^2 \tag{6.5}$$

$$\tilde{\kappa}_3 = \tilde{m}_3 - 3\tilde{m}_1\tilde{m}_2 + 2\tilde{m}_1^3 \tag{6.6}$$

Thanks to the relation of logarithmic cumulants to the moments of logarithms of u defined in Tison *et al.* [2004], we get the empirical expressions for the logarithmic cumulants as we earlier mentioned in the equation 6.3 as follows:

$$\tilde{\kappa}_1 = \tilde{m}_1 = \mathbb{E}[\ln u] \tag{6.7}$$

$$\tilde{\kappa}_2 = \tilde{m}_2 - \tilde{m}_1^2 = \mathbb{E}[(\ln u - \tilde{\kappa}_1)^2] \tag{6.8}$$

$$\tilde{\kappa}_3 = \tilde{m}_3 - 3\tilde{m}_1\tilde{m}_2 + 2\tilde{m}_1^3 = \mathbb{E}[(\ln u - \tilde{\kappa}_1)^3] \tag{6.9}$$

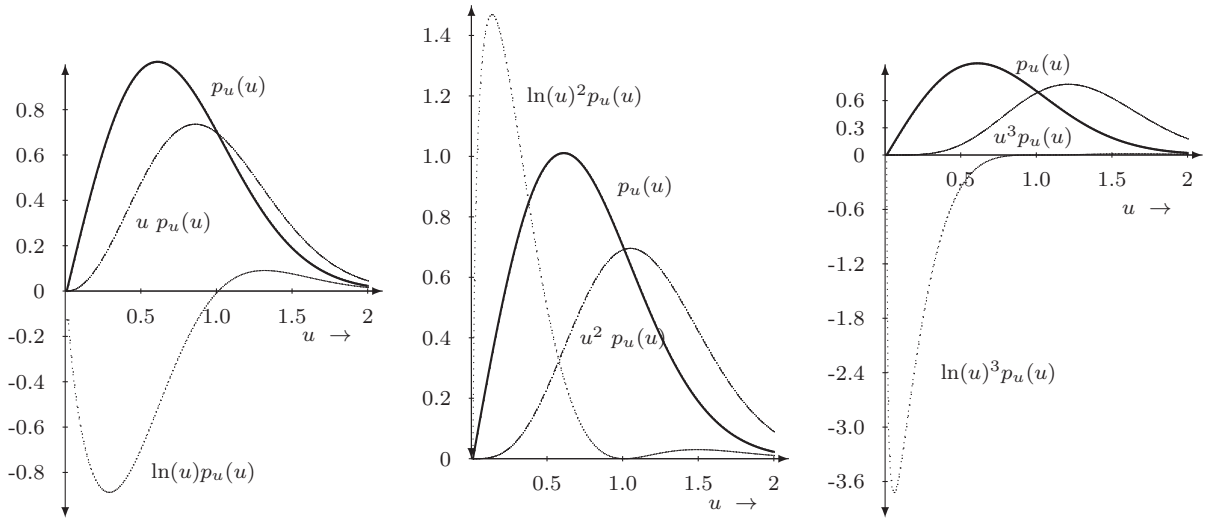
So in other words, the first three logarithmic-cumulants can be called logarithmic-mean, logarithmic-variance and logarithmic-skewness. The computation of first three sample logarithmic-cumulants is simply carried out by first taking the natural logarithm of the data and then computing the mean, variance and skewness. The expressions logarithmic-moments and logarithmic-cumulants and their definitions are based on the Mellin transform. The use of Mellin transform becomes possible in case of the SAR images as radar amplitude or intensity is intrinsically positive. Even for the complex-valued data, when we apply some linear transformations to achieve a multi-scale representation such as the FrFT (multi-scale in phase) or Gabor filter (oriented sub-bands), their respective amplitude values will also be intrinsically positive. So from the classical statistical understanding which the moment generating function given by the Fourier transform of the probability density function should be replaced by some transform defined only of the positive values, such as Mellin transform. We know that Fourier transform represents an integral over a set of all real values, where its counterpart in the $\mathbb{R}^+$ space, i.e. the Mellin transform provides a natural way to define the distribution of products of independent random variables (Tupin [2011]). Which is a more plausible formulation specifically for SAR images.

Now we discuss general properties of the moments, which will lead us to the demonstrating importance of the logarithmic-cumulants for SAR relevant feature descriptors over moments. For a probability density function $p_u(u)$ of set of random variables u, considering the case of moment generating function, the distribution is multiplied by u^i while computing the i^{th} order moment μ_i as follows:

$$\mu_i = \int_{-\infty}^{+\infty} u^i p_u(u) du. \tag{6.10}$$

So the importance of tail is more emphasized in higher moments when i increases. In such case the moments are not correct representation of the statistical response of the data as they are influenced by strong scatterers found at the tail of the distribution. So in case of the SAR images, where the targets of importance are not those strong scatterers which appear at the tail of the distribution, the moments are not the correct statistical representation of the data. While when using the Mellin transform instead, and computing the logarithmic-cumulants, both the tail and head of the distribution are taken into account. Thus the importance of Mellin transform is

even more justified given the fact that mostly the SAR images have heavy-tailed distribution. Lets demonstrate this concept with an example presented in the figure 6.11.



(a) First order. (b) Second order. (c) Third order.

Figure 6.11: Interest of logarithmic-cumulants for an improved feature descriptor. Solid line indicates the distribution, $p_u(u)$. Dashed line (- - -) indicates the computation of i^{th} moment, $u^i p_u(u)$. Dotted line $(\cdots)$ indicates the computation of i^{th} order logarithmic-cumulants, $(\ln\ u)^i p_u(u)$. First, second and third order cases are shown in (a), (b) and (c), respectively. In case of the moments, the distribution is multiplied by u^1, u^2, and u^3 for the first, second and third moments. So for higher moments tail plays deciding factor, which could be corrupted by strong scatterers. Whereas in the case of logarithmic-cumulants, first the logarithm of the data is computed and later over this moments are estimated. Its evident from the shown plots that both the head and tail of the distribution are taken into account, thus logarithmic-cumulants is a better statistical representation of the images for heavy-tailed distribution cases.

The figure demonstrates a Rayleigh distributed $p_u(u)$ for scale parameter 0.5. It can be seen from the figure 6.11 that for the computation of moments, the tail of the distribution is taken a lot into account. Tail of the distribution play an important role in statistical characterization for higher moments as here the distribution $p_u(u)$, the distribution is multiplied with u^i before integration in the computation of i^{th} moment. But this could lead to an erroneous statistical characterization in the heavy-tail case, which is very common in SAR images. Now the contribution of head of the distribution is enhanced in the statistical characterization with logarithmic-cumulants. Different logarithmic-cumulants have certain interpretations for the amplitude and intensity SAR images. The first logarithmic-cumulant (logarithmic-mean: κ_1) obviously measures the amount of backscatter, the second logarithmic-cumulant (logarithmic-variance: κ_2) can be related to the amount of texture (deviation from Gaussian statistics) and the third logarithmic-cumulant (logarithmic-skewness: κ_3) measures the type of texture (negative $\mathcal{K}$-distribution-like skew or positive $\mathcal{G}^0$-distribution-like skew, measure relative to the Gaussian scale). At least for a linear transformation such as the FrFT or Gabor filters, the interpretations of the logarithmic-cumulants of the transformed coefficients essentially remains the same.

6.4.3.2 Logarithmic cumulants as feature descriptor

Logarithmic cumulants for FrFT coefficients Let $\mathrm{FrFT}_\alpha(\cdot)$ be the 2-D fractional Fourier transform operator at a transform angle α. For an image $I(x, y)$ of size M $\times$ N, the first three

logarithmic-cumulants ($\tilde{\kappa}_{1_\alpha}$, $\tilde{\kappa}_{2_\alpha}$ and $\tilde{\kappa}_{3_\alpha}$) for the amplitude-envelope of the fractional Fourier transform coefficients at a transform angle α are estimated as follows:

$$\tilde{\kappa}_{1_\alpha} = \frac{1}{\mathrm{M}\cdot\mathrm{N}}\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}}\ln\left(|\mathrm{FrFT}_\alpha(I(x,y))|\right), \tag{6.11}$$

$$\tilde{\kappa}_{2_\alpha} = \frac{1}{(\mathrm{M}\cdot\mathrm{N}-1)}\times\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}}\left(\ln\left(|\mathrm{FrFT}_\alpha(I(x,y))|\right)-\tilde{\kappa}_{1_\alpha}\right)^2, \tag{6.12}$$

$$\tilde{\kappa}_{3_\alpha} = \frac{\mathrm{M}\cdot\mathrm{N}}{(\mathrm{M}\cdot\mathrm{N}-1)(\mathrm{M}\cdot\mathrm{N}-2)}\times\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}}\left(\ln\left(|\mathrm{FrFT}_\alpha(I(x,y))|\right)-\tilde{\kappa}_{1_\alpha}\right)^3, \tag{6.13}$$

with $\alpha = 0, \Delta\alpha, 2\Delta\alpha, 3\Delta\alpha, \cdots, \pi$. So for $\Delta\alpha = \pi/16$, the feature descriptor of the length 51 has the following form:

$$\bar{f}_{\mathrm{FrFT}\ \tilde{\kappa}} = \{\tilde{\kappa}_{1_0}, \tilde{\kappa}_{2_0}, \tilde{\kappa}_{3_0}, \tilde{\kappa}_{1_{\Delta\alpha}}, \tilde{\kappa}_{2_{\Delta\alpha}}, \tilde{\kappa}_{3_{\Delta\alpha}}, \cdots, \tilde{\kappa}_{1_\pi}, \tilde{\kappa}_{2_\pi}, \tilde{\kappa}_{3_\pi}\}. \tag{6.14}$$

Logarithmic cumulants for Gabor filtered images Let $\mathrm{Gabor}_{\mathrm{S,O}}(\cdot)$ be the Gabor filtered image at a stage **S**=S, and orientation **O**=O. For an image $I(x,y)$ of size $\mathrm{M}\times\mathrm{N}$, the first three logarithmic-cumulants ($\tilde{\kappa}_{1_\alpha}$, $\tilde{\kappa}_{2_\alpha}$ and $\tilde{\kappa}_{3_\alpha}$) for the amplitude-envelope of the Gabor filtered image at at a stage **S**=S, and orientation **O**=O are estimated as follows:

$$\tilde{\kappa}_{1_{\mathrm{S,O}}} = \frac{1}{\mathrm{M}\cdot\mathrm{N}}\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}}\ln\left(|\mathrm{Gabor}_{\mathrm{S,O}}(I(x,y))|\right), \tag{6.15}$$

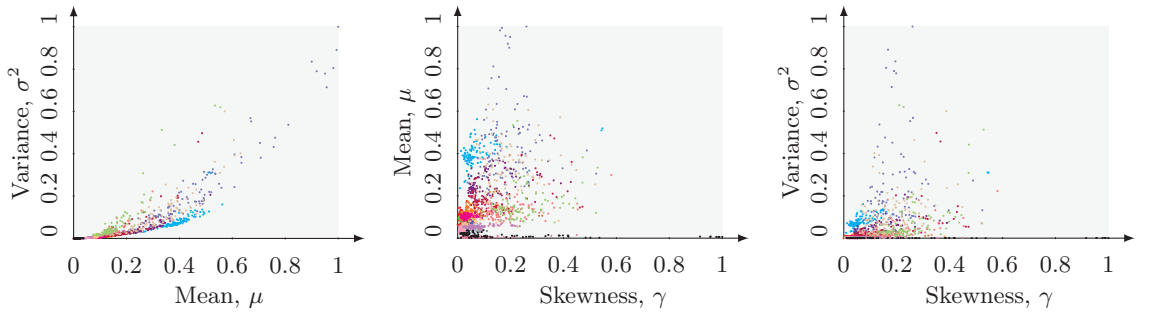
$$\tilde{\kappa}_{2_{\mathrm{S,O}}} = \frac{1}{(\mathrm{M}\cdot\mathrm{N}-1)}\times\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}}\left(\ln\left(|\mathrm{Gabor}_{\mathrm{S,O}}(I(x,y))|\right)-\tilde{\kappa}_{1_{\mathrm{S,O}}}\right)^2, \tag{6.16}$$

$$\tilde{\kappa}_{3_{\mathrm{S,O}}} = \frac{\mathrm{M}\cdot\mathrm{N}}{(\mathrm{M}\cdot\mathrm{N}-1)(\mathrm{M}\cdot\mathrm{N}-2)}\times\sum_{u=1}^{\mathrm{M}}\sum_{v=1}^{\mathrm{N}}\left(\ln\left(|\mathrm{Gabor}_{\mathrm{S,O}}(I(x,y))|\right)-\tilde{\kappa}_{1_{\mathrm{S,O}}}\right)^3, \tag{6.17}$$

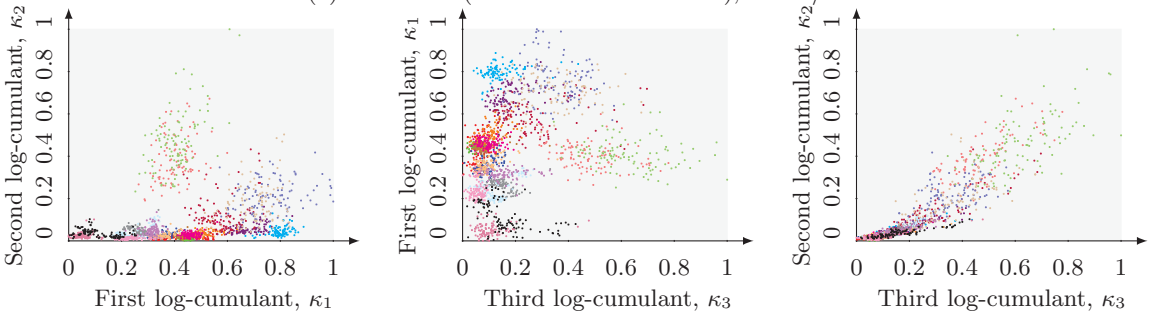
Thus, for 3 stages, and 4 orientation, we obtain a feature descriptor of length 36 as follows:

$$\begin{aligned}\bar{f}_{\mathrm{Gabor}\ \tilde{\kappa}} = \{&\tilde{\kappa}_{1_{1,1}}, \tilde{\kappa}_{2_{1,1}}, \tilde{\kappa}_{3_{1,1}}, \tilde{\kappa}_{1_{1,2}}, \tilde{\kappa}_{2_{1,2}}, \tilde{\kappa}_{3_{1,2}}, \tilde{\kappa}_{1_{1,3}}, \tilde{\kappa}_{2_{1,3}}, \tilde{\kappa}_{3_{1,3}}, \tilde{\kappa}_{1_{1,4}}, \tilde{\kappa}_{2_{1,4}}, \tilde{\kappa}_{3_{1,4}},\\ &\ldots\tilde{\kappa}_{1_{3,1}}, \tilde{\kappa}_{2_{3,1}}, \tilde{\kappa}_{3_{3,1}}, \tilde{\kappa}_{1_{3,2}}, \tilde{\kappa}_{2_{3,2}}, \tilde{\kappa}_{3_{3,2}}, \tilde{\kappa}_{1_{3,3}}, \tilde{\kappa}_{2_{3,3}}, \tilde{\kappa}_{3_{3,3}}, \tilde{\kappa}_{1_{3,4}}, \tilde{\kappa}_{2_{3,4}}, \tilde{\kappa}_{3_{3,4}}\}.\end{aligned} \tag{6.18}$$

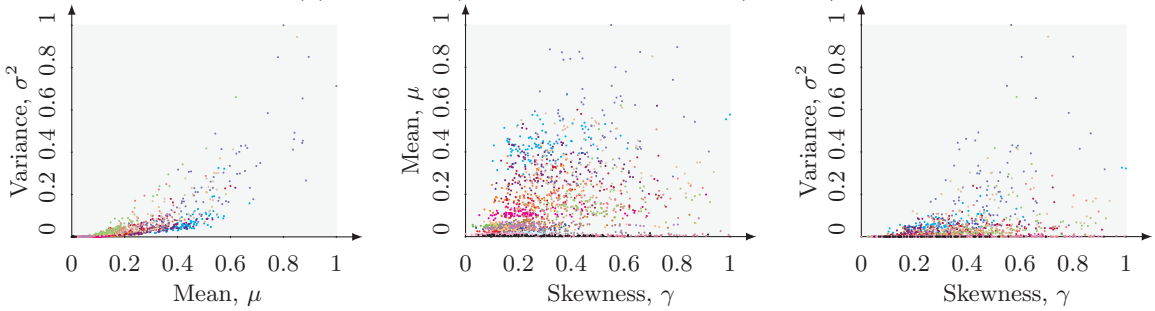
Before carrying out the mean and individual accuracy analysis of proposed logarithmic-cumulants based method and earlier discussed well performing methods, we have a look at the feature space of the logarithmic-cumulants based feature descriptor in comparison to the moments based feature descriptor for the FrFT and Gabor case. It can be seen in the figure 6.12, that the category separability is enhanced in the feature space of the logarithmic-cumulants over the moments feature space. In the further discussions, the feature descriptor for FrFT and Gabor based on logarithmic cumulants will be denoted as $\mathbf{FrFT}_\kappa$ and $\mathbf{Gabor}_\kappa$, respectively.



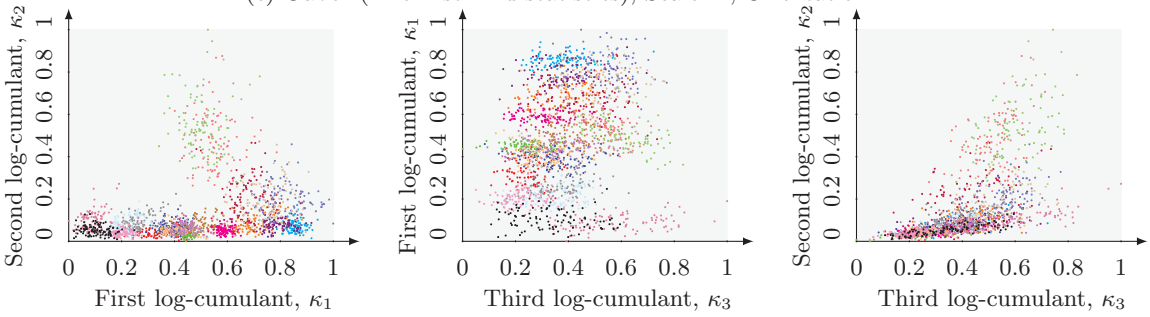
(a) The FrFT (The first kind statistics), $\alpha = \pi/8$



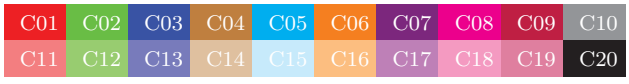
(b) The FrFT (The second kind statistics), $\alpha = \pi/8$



(c) Gabor (The first kind statistics), Scale: 1, Orientation: 2



(d) Gabor (The second kind statistics), Scale: 1, Orientation: 2



Category color-index

Figure 6.12: Feature space diagram demonstrating the potential of the considered nonparametric feature descriptors while using the logarithmic-cumulants for category separability. Selected 2-D planes in the feature space for corresponding feature descriptor is presented, however, it must be noted that the actual feature space is multi-dimensional. The dimensionality of feature space is same as the length of respective feature descriptor. Some feature have very large positive or negative numerical values, thus for a better interpretation, all the features have been scaled between 0 and 1.

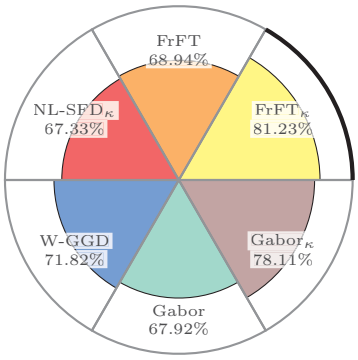
6.4.4 Detailed assessment of individual category accuracies

On a methodological point of view, we can keep **W-GGD**, and **Gabor**/**Gabor**$_\kappa$ in one group, and **NL-SFD**, and **FrFT**/**FrFT**$_\kappa$ in another. The first group contains methods which are multi-scale / multi-resolution methods. These methods are very successful in texture synthesis, and even for very intricate arrangement of objects in an image patch. The second group corresponds to methods based directly on Fourier transform or generalization of Fourier transform. These methods are not only capturing the texture information but more importantly the information on the backscattering behavior of the objects on ground. Actually second group must be termed as a true multi-scale approach in case of SAR images, where the scaling is carried out in the phase.

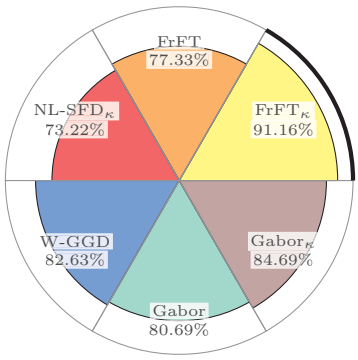
Analyzing the first polar chart in the figure 6.13 for mean accuracy, we notice that the logarithmic-cumulants based **FrFT**$_\kappa$ provides best accuracy of 81.23%, improvement of about 12% over moments based **FrFT** method, and about 14% better than **NL-SFD**. The second best mean accuracy could be achieved by logarithmic-cumulants based **Gabor**$_\kappa$, improvement of about 10% over moments based **Gabor** method, and about 06% better than **W-GGD**. The moments (mean: μ, variance: σ^2, and skewness: γ) and logarithmic cumulants (logarithmic-mean: κ_1, logarithmic-variance: κ_2, and logarithmic-skewness: κ_3 or simply logarithmic-cumulants: κ) are sometimes referred as 'first kind statistics' and the 'second kind statistics', respectively. The highest accuracy has been highlighted with an bold arc on the outer circle representing 100% accuracy level for respective feature descriptors in the polar charts in the figure 6.13. The individual category accuracy assessment in these polar charts suggest that best individual accuracy is achieved either in **FrFT**/ **FrFT**$_\kappa$ or **Gabor**$_\kappa$ based methods. Thus it is sufficient to discuss these two methodologies in detail, as done in following.

6.4.4.1 On the accuracy of the Gabor/Gabor$_\kappa$ based feature descriptors

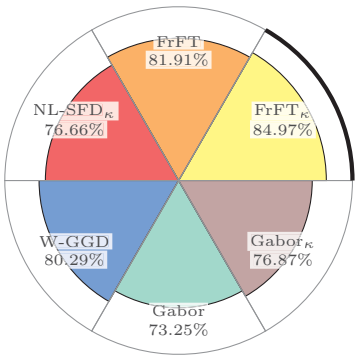
In many image retrieval systems, Gabor filter-bank based texture feature extraction algorithms measure the weighted energy of filtered images, regarded as feature descriptors, assuming that the energy distribution within the frequency domain describes the texture. Alternatively, the 'first kind statistics (moments)' of filtered images can be considered as a texture measure. The most essential property of Gabor coefficients for characterizing the texture in images is its bandpass filter property, which has been demonstrated in Farrokhnia & Jain [1991]. This is also the underlying mechanism in humans' visual system, which we try to imitate in the Gabor expansion of signals to understand the texture. The homogeneous texture descriptor under the visual texture descriptor in the moving picture experts group (MPEG)-7 standard (Sikora [2001]) also relies on the moments of Gabor filtered images, and can be considered as a benchmarking tool for comparing alternative methodologies. The Gabor wavelets used in Gabor filter banks are non-orthogonal in nature, unless they are poorly concentrated in time or frequency. While keeping a good time and frequency domain concentration, non-orthogonality is allowed at the cost of keeping redundant information in the filtered image. To reduce this redundancy, the Gabor kernel parameter should be carefully selected (Manjunath & Ma [1996]). The results in the figure 6.13 suggest that the use of logarithmic-cumulants of the Gabor filtered images outperform traditional strategies (i.e. the moments of Gabor filtered images) with substantial improvement in accuracy in majority of categories. Overall mean accuracy enhancement is about about 10%.



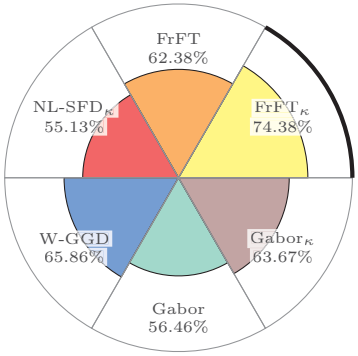
(a) Mean Accuracy



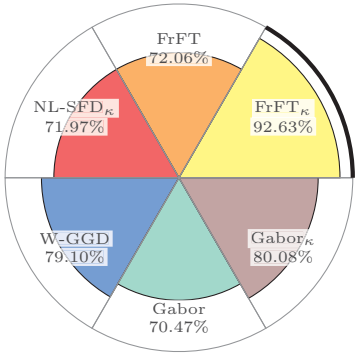
(b) Category 01



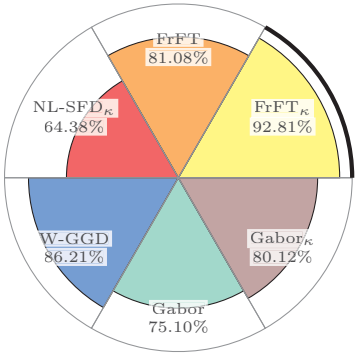
(c) Category 02



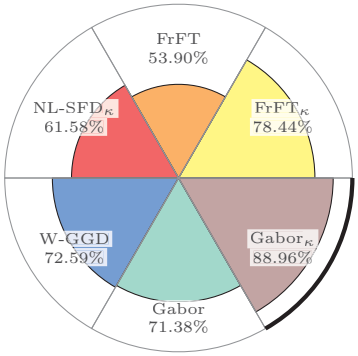
(d) Category 03



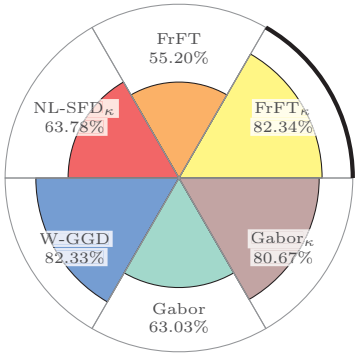
(e) Category 04



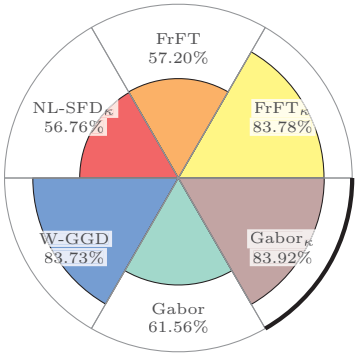
(f) Category 05



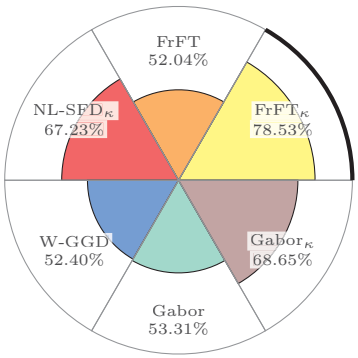
(g) Category 06



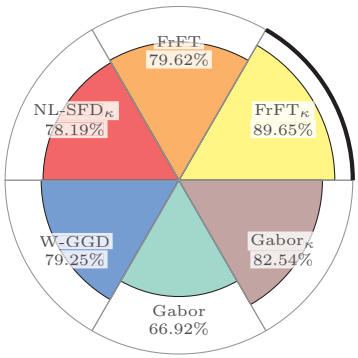
(h) Category 07



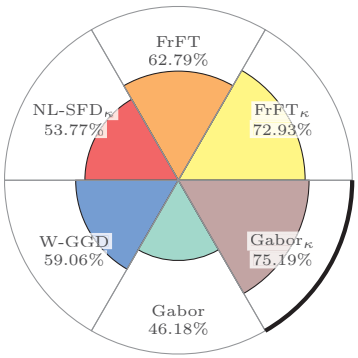
(i) Category 08



(j) Category 09

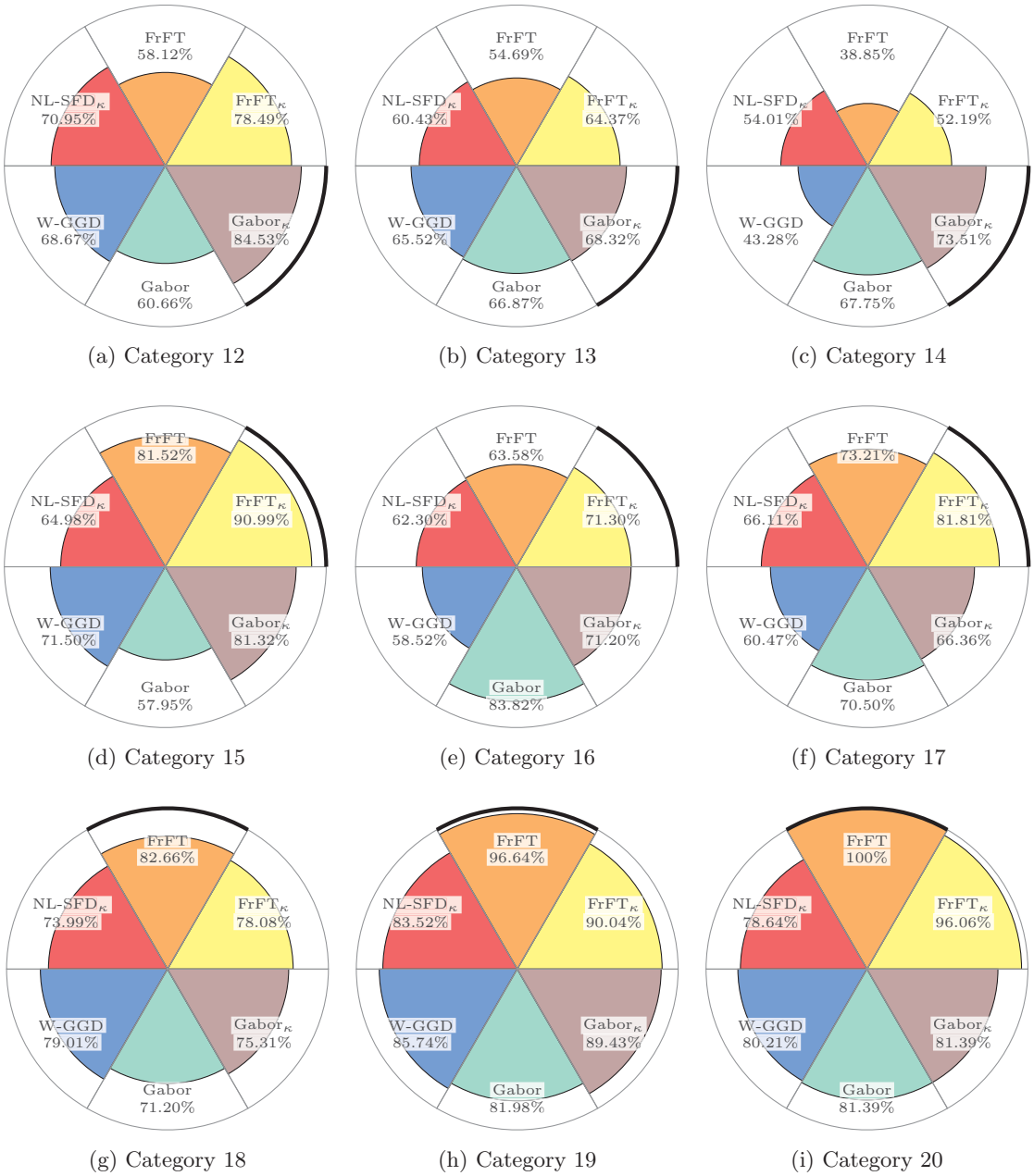


(k) Category 10



(l) Category 11

contd.



(a) Category 12 (b) Category 13 (c) Category 14

(d) Category 15 (e) Category 16 (f) Category 17

(g) Category 18 (h) Category 19 (i) Category 20

Figure 6.13: Improvement in the accuracy of patch-oriented image categorization with the use of logarithmic cumulants ($\kappa's$) for the amplitude envelope of the fractional Fourier transform coefficients and Gabor filtered images. The feature descriptor with highest accuracy in each category has been highlighted with an bold arc on the outer circle representing 100% accuracy level. First polar chart on the previous page represents the overall mean accuracy. Overall accuracy assessment suggests that **FrFT**$_\kappa$ provides best overall mean accuracy, followed by **Gabor**$_\kappa$. The accuracies of all other methods are more or less similar. However, here again individual accuracy assessment suggests that no feature descriptor outperforms all other in each individual category scenario, thus makes it difficult to arrive at a general consensus on a best performing method. But the deeper analysis of image content in individual categories will explain the performance of the respective feature descriptor. This knowledge will aid us to decide and select a particular feature descriptor depending upon the desired application.

6.4.4.2 On the accuracy of the FrFT/FrFT$_\kappa$ based feature descriptors

Unlike Gabor-based methods, where the texture of the images plays an important role, the fractional Fourier transform based methods mainly discriminate the images based on the underlying backscattering phenomenon on the ground. The fractional Fourier transform is the projection of the signal on the rotated time-frequency plane, which essentially uses the phase information embedded into the complex-valued SAR images. The signal at each plane, i.e. at each transform angle, represents its decomposition in terms of chirps. Here the frequency modulation rate (FM rate) of the chirp is determined by the transform angle. As the FrFT based method uses the complex-valued data, the full information (amplitude and phase) is exploited while forming sub-bands at different transform angles (α's). The FrFT is actually the true multi-scale transform for the SAR images, where the scaling is carried out in the phase, and scale factor is controlled by the transform angle α. The FrFT can be viewed as a Radon transform on the time-frequency Wigner-Ville distribution (Cohen [1989]), i.e. WV marginal distribution on a straight line parametrized by α in the time-frequency plane. So the WVD rotated by α in the time-frequency plane is essentially equivalent to the FrFT.

The use of logarithmic-cumulants for the FrFT coefficients performs better than the moments of the FrFT coefficients based feature descriptor in the majority of the categories with a higher dynamic range. The logarithmic-transformation induces symmetry by reducing the dynamic range. However the categories, C18-*Sandy plains*, C19-*Buoy in Sea* and C20-*Ambiguities on the sea surface* seem to perform better for moments case. It is very important to understand these results for which we refer again to the figure 6.11. We noticed there that in the estimation of the logarithmic-cumulants, both the head and tail of the distribution is taken into account, whereas in the estimation of the moments, the importance of tail of the distribution is understandable. Now in these three categories, the targets which actually describes the content of the image lies on the tail of the distribution. E.g. in category C19-*Buoy in Sea*, the strong scattering buoys are essentially at the tail of the distribution which is the actual information content in this category. As well as in category C20-*Ambiguities on the sea surface*, the ambiguities from the nearby building have stronger response. Similarly, in category C18-*Sandy plains*, where there is a lot of clutter due to sandy surface, only the small strong scatterers are actually differentiating it from other smooth surfaces with diffused scattering such as C01-*Grassland*.

6.4.4.3 The FrFT-based feature descriptors versus the Gabor-based feature descriptors

The Gabor filters are normally more suitable for strongly textured images, however the FrFT coefficients captured the signals' behavior efficiently relevant to the SAR images. In this section, we will provide reasoning for poor accuracy for each category, in order to better understand the strengths and weaknesses of the FrFT-based (**FrFT/FrFT**$_\kappa$) and Gabor-based (**Gabor/Gabor**$_\kappa$) methods in general.

The texture in the categories C13-*Skyscrapers* and C14-*Industrial complex* is quite distinct, but the backscattering phenomena on the ground for both is very similar, as it is characterized by multiple bounces from very-strong scatterers. That's why the FrFT-based method groups many patches into the same category, leading to a lower accuracy for C13-*Skyscrapers* and C14-*Industrial complex*. Similar is the case for categories C11-*Urban coastline* and C12-*Aircraft stands*. Both categories contain regions with strong scatterers (urban structures in C11 and structures of aircraft-stands in C12) and specular reflection regions (calm water in C11 and concrete surface in C12).

In the case of C06-*Urban-02*, and C08-*Urban-04*, the texture is very distinct, however the

scattering mechanism is similar due to the similar type of material used in constructing building-blocks, as well as due to the presence of vegetation in-between. Thus, the Gabor-based method performs better than the FrFT-based methods. The situation is different in C05-*Urban-01*, and C07-*Urban-03*, in which the backscattering behavior is different whereas the texture is very much the same; thus FrFT-based method outperforms the Gabor-based methods.

The uniform textured land-covers, i.e. C01-*Grassland*, C02-*Forest*, and C18-*Sandy plains* are grouping with each other in case of the Gabor-based method, whereas the FrFT-based method is able to better discriminate between them. The similarity in the texture for the categories C04-*Vegetation and a water-channel*, C10-*Natural water-channel*, C15-*Airstrip*, and C17-*Train tracks* also leads to false-alarms if Gabor-based methods are applied. In spite of the similarity in texture between C04, C10, C15, and, C17, these categories exhibit different backscattering properties due to different materials on the ground, therefore the FrFT-based method is more suitable. The category C09-*Urban water-channel* also contains false-alarms from C07-*Urban-03* for the case of Gabor-based methods.

The last two categories, C19-*Buoy in Sea* and C20-*Ambiguities on the sea surface*, are the categories with very distinct signatures visible only in the fractional Fourier transform domain. Thus the FrFT-based methods outperforms Gabor-based methods.

Summary For the patch-oriented image categorization, the following conclusions can be drawn with the demonstrations on experimental database used, in context to the methodological framework presented in the figure 6.1:

- The model parameter methods applied in Bayesian framework are not found to be very much suitable for the patch-oriented image categorization. The reason is increase in heterogeneity in a larger patch leads to an inaccuracy in the model parameter estimation. It is not possible to describe a larger patch with diversity of objects with few model parameters.

- In general, methods applied on the image in a suitable transformation space, provide better accuracy. The reason is that in various sub-bands some information about the objects and intrinsic texture is revealed which is otherwise hidden in spatial domain representation of the image.

- A newly proposed method based on the fractional Fourier transform essentially proven to be a true multi scale approach. This method captures the important primitive feature, i.e. the backscattering behavior on ground better than the wavelet-based parametric method and Gabor-based nonparametric methods which are suitable for very well textured images.

- Apart from the standard moments based feature descriptor, the use of logarithmic-cumulants provides substantially enhanced overall accuracy. The logarithmic transformation balances the influence of head and tail of the distribution leading to a better statistical characterization of the distribution. However, in FrFT-based methods, the categories where the target of interest essentially lies on the tail of the distribution, the moments should be preferred over the logarithmic-cumulants.

6.5 Summary

The chapter provides critical evaluation of the traditional, state-of-the-art and proposed methods to generate a primitive feature descriptor, in context to the methodological framework presented in the figure 6.1. Following is the summary of the results presented in this chapter:

- **Pixel-based classification**

  Gibbs based parametric approaches are capable of discriminating the textures even at the smaller window sizes of 16×16 pixels for pixel-based classification. These methods are certainly useful for the land-cover mapping and monitoring of the regions with underlying homogeneous textures. Although such regions can be found in high and very-high-resolution SAR images from satellites such as TerraSAR-X, TanDEM-X etc., but this will find more applications in upcoming lower-resolution satellites such as Sentinel-1 for large scale mapping. The only limitation is the computational cost in texture model parameter estimation in Gibbs based methods. This should be overcome with efficient implementation of the algorithms and improved hardware.

  With larger analyzing window sizes of 64×64 pixels, the methods applied to the images in transformation domain also start to perform better in comparison to the lower analyzing window of 16×16 pixels. But further increasing the analyzing window will lead us to a risk of heterogeneity in the texture, which should be rather considered as a case of patch-oriented image categorization.

- **Patch-oriented image categorization**

  For the patch-oriented image categorization, the Gibbs based methods are not as successful as for the pixel-based classification. The approaches applied on the image in transformation domain, be they parametric or nonparametric, perform very good.

  The proposed use of logarithmic-cumulants essentially enhances the accuracy in the sub-bands based methods (i.e. for the FrFT and Gabor filter banks) traditionally using the moments as simple statistical measures of the sub-bands. The reason behind this improvement is that logarithmic-cumulants emphasizes the contribution of the head and tail of the distribution, whereas the tail of the distribution is taken into account a lot in the computation of the higher moments. The presence of strong scatterers at the tail of the distribution, which actually might not be the true representation of the image leads to inaccurate statistical characterization with moments. However, it must be noted that the images where the most important targets are the strong scatterers, the moments are still the better choice.

  Although no single feature descriptor performs the best in all categories, but the proposed fractional Fourier transform based methods outperforms all other methods in terms of overall mean accuracy.

In summary, we can again comment that it is difficult to arrive a general consensus over a method performing best in all the scenarios. But in general, the Gibbs based parametric methods are certainly powerful tool for pixel-based classification. Feature descriptors computed on images in transformation domain should be preferred for the patch-oriented image categorization. The presented study, experimental database and critical evaluation can be considered as a benchmark while developing further algorithms for various applications for spatial content understanding in SAR images.

Chapter 7

Summary and conclusions

This thesis analyzes both existing and newly proposed algorithms for the primitive feature extraction in synthetic aperture radar (SAR) images. We propose a methodological classification framework (Singh & Datcu [2012a]) for the qualitative and quantitative evaluation and comparison of the studied methods. This study was motivated by the availability of metric-resolution SAR images from last generation satellites such as TerraSAR-X, TanDEM-X, COSMO-SkyMed, Radarsat-2 etc., where the information available in a resolution cell is large enough to carry out individual target analysis. The statistical characterization of such SAR images also demands new approaches.

The earlier statistical algorithms were based on the understanding of speckle characteristics in SAR images. The focus has been generally on the model-based textural parameter estimation in the amplitude SAR images, such as parametric Gibbs-based methods in the Bayesian framework. Two Gibbs-based methods, namely, the Gauss-Markov random fields (GMRF) and the autobinomial model (ABM), have been evaluated in this thesis. In certain cases, the assumption of fully developed speckle, necessary for applying Gibbs-based methods to SAR data, is not fulfilled in some or all of the regions in the image. In order to handle this problem, we suggest to evaluate also nonparametric approaches which do not make any assumption on the adapted data model.

To be able to understand the content of high and very-high-resolution SAR images, we presented some examples of the metric-resolution spotlight mode TerraSAR-X images and compared them with the low resolution ERS images. In certain scenarios, where the homogeneous texture is visible, it is still possible to carry out pixel-based classification with parametric methods, even in high-resolution SAR images. This can be applied only for small scale mapping and monitoring, as the swath width reduces substantially with the increase in resolution. But when the details available in a small region of the image too large, the pixel-based classification loses its advantages. In that case we proposed the patch-oriented image categorization, meaning that our objective is not to classify some regions of the image, but the whole image-patch into an object category.

For practical applications, we need to generate a compact feature descriptor which is defined with a set of primitive features describing the spatial content of the image. The results presented in this thesis demonstrated the importance of a SAR-relevant primitive feature, termed as the 'backscattering response'. Traditional primitive features such 'texture, shape and color' which are generally used to characterize an image are found insufficient especially for the case of very-high-resolution SAR images. The information on the 'backscattering response' is better understood with analysis of the complex-valued SAR images, rather than using just its amplitude information. We demonstrated this notion with a proposed methodology for the visual

exploration of individual targets in complex-valued SAR images based on the concept of azimuth splitting. This methodology, termed as multiple sublook decomposition approach (MSLD Singh & Datcu [2012c]) generates hyper-images, utilizing the phase information, which is not available in the amplitude SAR images. The MSLD motivated us to consider the complex-valued SAR images for information extraction, the use of which is generally limited to interferometric or polarimetric applications only.

The exploitation of the phase information in complex-valued SAR image can be carried out using a chirplet-derived linear transformation, e.g. the fractional Fourier transform (FrFT). In literature, the use of the FrFT is restricted to moving targets detection. The analysis and experiments suggested the suitability of the FrFT for a broader spatial content understanding in SAR images (Singh & Datcu [2012b]). The advantage of the FrFT is its multi-scaling property, where the scaling is done in the phase. Now the task even in the case of FrFT-based method is to generate a compact feature descriptor. The general choice will be to compute simple statistical measure such as moments up to certain order on the amplitude-envelope of the FrFT coefficients. We suggested to enhance the performance of this method by introducing the use of logarithmic-cumulants as an alternative to the moments (Singh & Datcu [2013]). The distribution of the amplitude of the FrFT-coefficients is heavy-tailed like the original SAR images, so the moments are not providing the correct statistical characterization of the distribution of the coefficients. The reason is the importance of tail of the distribution in the higher moments, which may contain many strong scatterers which are not essential for describing the spatial content of the data. On the contrary, the logarithmic-cumulants provide a balanced contribution of the head as well as the tail of the distribution, which in turn seems to better represent the spatial content of the image under consideration.

Summary of results and conclusions

The evaluation of various techniques has been carried out under a proposed robust methodological classification framework, motivated by increase in the diversity of the targets and objects in the high and very-high-resolution SAR images. In order to evaluate the performance of the considered parametric and nonparametric approaches, we used an experimental database generated from metric-resolution spotlight mode TerraSAR-X images.

For pixel-based classification, the database includes 12 sub-images of size 256×256 pixels containing distinctive homogeneous texture regions, ranging from various vegetation types, urban areas to calm and rough water. The k-nearest neighbors (k-NN) supervised classification has been carried out with 1% and 10% training samples, with an analyzing window (aw) of size 16×16, 32×32, and 64×64 pixels. Considering the case of the 64×64 pixels analyzing window, at 10% training sample size, there is no significant difference in the accuracy of parametric and nonparametric approaches. At a smaller analyzing window (16×16 pixels), it is possible to achieve good accuracy with Gibbs-based parametric approaches applied to the image domain. These methods are therefore useful for the land-cover mapping and monitoring of regions with underlying homogeneous textures. The limitation of the computational cost in Gibbs based methods can be overcome via an efficient implementation of the algorithms and by improving the hardware.

For patch-oriented image categorization, the database constituted 2000 image patches from 20 well-defined categories, with 100 patches in each category. The k-NN supervised classification has been carried out with 10% training samples. In general, approaches applied on the images within a transformation domain, parametric as well as nonparametric, outperformed the others. A substantial accuracy enhancement has been noticed when using logarithmic-cumulants of the

Gabor filtered images, instead of the moments. The proposed method of logarithmic-cumulant of the FrFT coefficients outperforms other traditional and state-of-the-art methods in overall accuracy. The presented results on the FrFT-based techniques motivate us to dwell into the analysis of complex-valued SAR images, in which the valuable phase information is present. The MSLD method of visual exploration of targets from the complex-valued SAR images further confirms our assumptions on the utility of the phase information in future algorithm developments.

Outlook and future work

The motivation to use complex-valued SAR images, the use of which was usually limited to interferometric and polarimetric applications, has been demonstrated in this thesis. A critical analysis of Gibbs-based parametric methods has also been carried out in this thesis. It is difficult to achieve a general consensus on best methods, as in our assessment no algorithm outperforms all others for every considered scenario. Thus, in the future, we propose to employ an automatic method selection technique, so that the parametric methods and nonparametric methods can be used in synergy, for practical applications with an enhanced overall accuracy.

The evaluation presented in this thesis provides a baseline study for parametric and nonparametric approaches. The selection of the most suitable method depends on the applications specific requirements. The proposed use of complex-valued SAR images is the underlying contribution of this thesis. This has been motivated by successful experiments carried out using MSLD and FrFT-based methods.

Appendices

Appendix A

TerraSAR-X

TerraSAR-X, the first German national remote sensing satellite in a public-private partnership between German Aerospace Center(DLR) and EADS Astrium Gmbh, was launched in June 2006 on a Russian/Ukrainian DNEPR-1 launch vehicle from the Baikonur cosmodrome in Kazakhstan. The objective of this satellite is to provide high quality radar image data to satisfy the data requirement for scientific research as well as data needs for commercial applications. It is the first commercial satellite in the world providing SAR images with metric resolution.

Figure A.1: TerraSAR-X

The satellite

The satellite, approximately weighing 1.3-ton, is based on the Flexbus-concept from EADS. This Flexbus-concept, which was also used successfully in CHAMP and GRACE missions before, greatly reduces the cost in comparison to conventional design. The satellite has a structure of hexagonal cross section with 5m long and 2.4m wide satellite bus, with one of the six sides carrying the 5m long and 80cm wide radar antenna. The enery requirements of the satellite are met by means of gallium arsenide solar cells. The downlink antenna which transfer the data recorded by radar antenna is secured to a 3.3m long mast in order to prevent the interference from radar antenna. Thus the acquisition of data and transmission of the data to ground station can be done simultaneously. On board GPS receiver is used for precise determination of the satellite orbit and star sensor ensures the antenna alignment accuracy as high as 65 arcsecs. Alongside the SAR instrument as primary payload, The Laser Communication Terminal (LCT) and The Tracking, Occultation and Ranging Experiment (TOR) are also flying on TerraSAR-X as secondary payloads.

SAR system and imaging modes

Primary payload, the SAR system has an active antenna which can be utilized in different modes. TerraSAR-X is operated in the X-band (frequency of around 9.65GHz, corresponding to the wavelength of about 3cm). The radar antenna which is 4.80m long and 80cm wide can operate in two polarizations, H (Horizontal) and V (Vertical). TerraSAR-X operates in 4 different modes depending upon the user requirements with following technical specifications:

High resolution spotlight mode Scene extension - 5km (azimuth) × 10 km (ground range), ground range resolution 1.48m - 3.49m (for incidence angle of 55° - 20°), azimuth resolution 1.1m(single polarization) and 2.2m (dual polarization), 91 full performance and 122 data access elevation beams, azimuth steering angle up to ± 0.75 °, possible polarizations HH or VV (single polarization) and HH/VV (dual polarization).

Spotlight mode Scene extension - 10km (azimuth) × 10 km (ground range), ground range resolution 1.48m - 3.49m (for incidence angle of 55° - 20°), azimuth resolution 1.7m (single polarization) and 3.4m (dual polarization), 91 full performance and 122 data access elevation beams, azimuth steering angle up to ±0.75°, possible polarizations HH or VV (single polarization) and HH/VV (dual polarization).

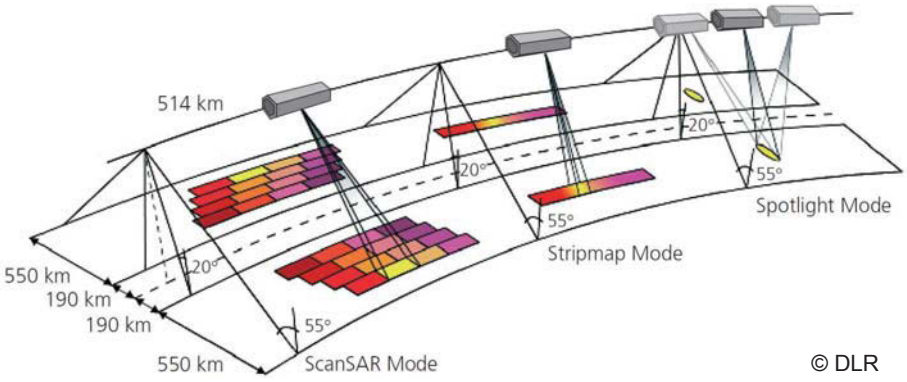


Figure A.2: TerraSAR-X imaging modes

Stripmap mode Swath width (ground range) of 30km with 4 sub-swaths, ground range resolution 1.70m - 3.49m (for incidence angle of 45° - 20°), azimuth resolution 3.3m (single polarization) and 6.6m (dual polarization), 12 full performance and 27 data access elevation beams, possible polarizations HH or VV (single polarization) and HH/VV, HH/HV, VV/VH (dual polarization).

ScanSAR Mode Swath width (ground range) of 100km with 4 sub-swaths, ground range resolution 1.70m - 3.49m (for incidence angle of 45° - 20°), azimuth resolution 1.85m, 27 full performance elevation beams.

High resolution spotlight data

Figure A.3 depicts the spotlight mode imaging in SAR. The phased array beam is steered in azimuth direction in order to increase the illumination time, thus increasing the size of syn-

thetic aperture CAF-DLR [2009]. A better azimuth resolution is obtained at the expense of reduced azimuth scene size. TerraSAR-X products are available in two different ranges of az-

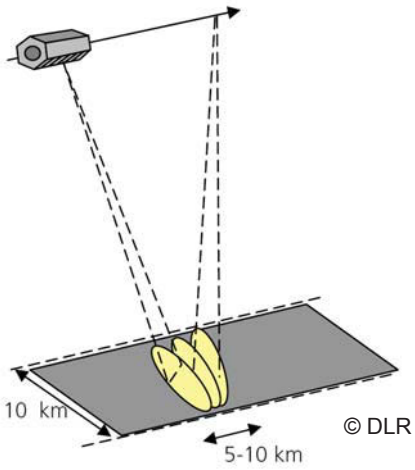


Figure A.3: Spotlight mode imaging

imuth resolution and azimuth scene size for spotlight mode, named as *Spotlight* (SL) and *High Resolution Spotlight* (HS). We have used HS data for our experimentation because it provides the best possible resolution from TerraSAR-X. The detailed characteristic values of HS mode are as follows CAF-DLR [2009]:

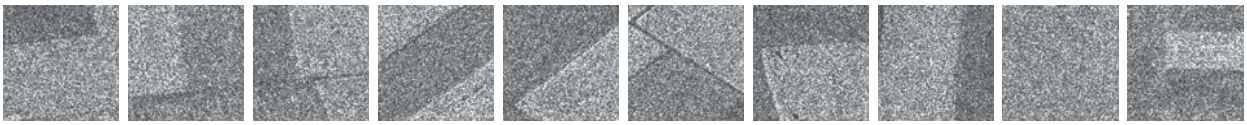
Table A.1: Parameters of high resolution spotlight mode

Parameter	Value
Scene Extension	5km (azimuth) × 10km (ground range)
Full performance incidence angle range	20° - 55°
data access incidence angle range	15° - 60°
Number of elevation beams	91 (full performance), 122 (data access)
Number of azimuth beams	up to 125 out of 229
Azimuth steering angle	up to ± 0.75°
Azimuth resolution	1.1m (single polarization)
	2.2m (dual polarization)
Ground range resolution	1.48m - 3.49m (at 55° - 20° incidence angle)
	0.74m - 1.77m (with 300 MHz bandwidth option
	and reduced swath extent in range)
Polarizations	HH or VV (single), HH/VV (dual)

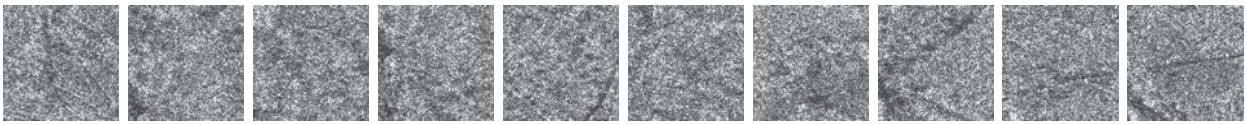
Appendix B

Experimental data base

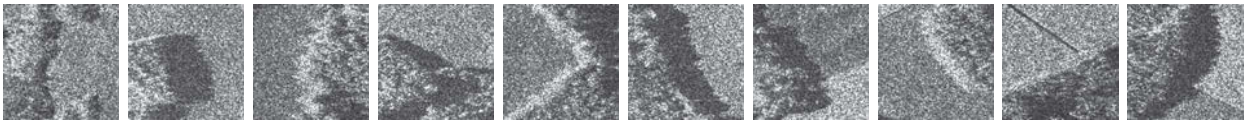
For the evaluation of feature descriptor performance in patch-oriented image categorization, a database of 2000 image-patches from the metric-resolution TerraSAR-X images has been generated. The size of each image patch is 200×200 pixels, thus each patch covers approximately 200^2 m^2 of ground. The considered experimental data base comprises of 20 well-defined objects / texture categories with 100 patches of each category. Example of one image-patch has been shown in the figure 6.7. Care has been taken to generate the database with diverse textures and objects in each category. In order to demonstrate this diversity, we present here 10 arbitrary selected image-patches from each category. The geographical locations from which each category originates is also presented, describing that the image-patches are from all over the world to include diversity and make them independent of local criteria.



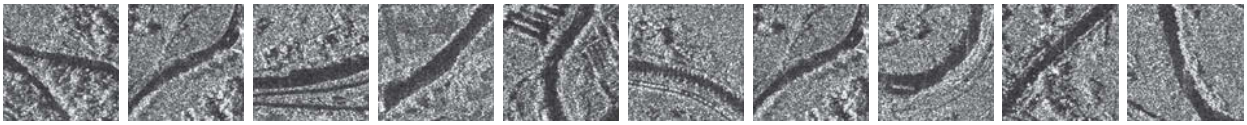
(a) **C01- Grassland.** Gilching and Munich (Germany), Stockholm (Sweden).



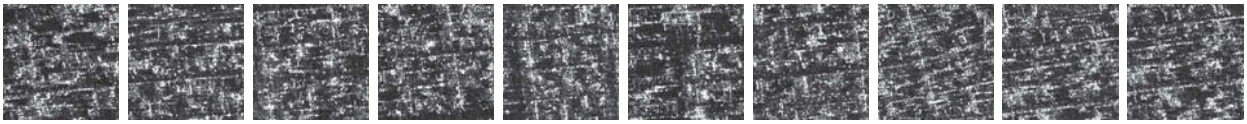
(b) **C02- Forest.** Paris (France), Gilching (Germany).



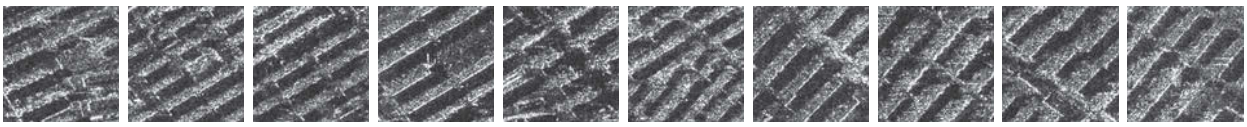
(c) **C03- Mixed vegetation.** Gilching (Germany), Stockholm (Sweden).



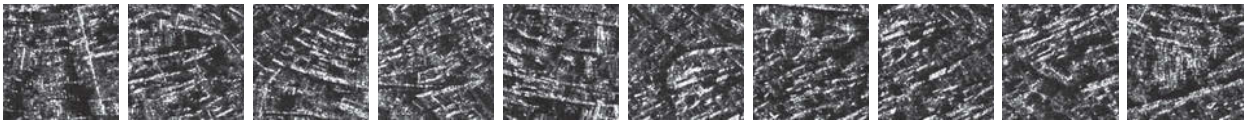
(d) **C04- Vegetation and urban with a water-channel.** Maribor (Slovania).



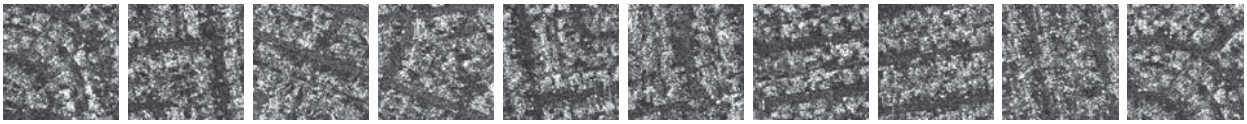
(a) **C05- Urban-01.** Tokyo (Japan).



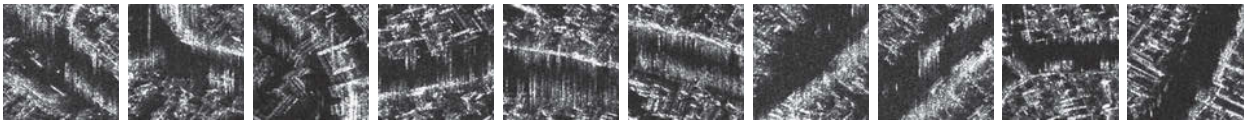
(b) **C06- Urban-02.** Munich (Germany).



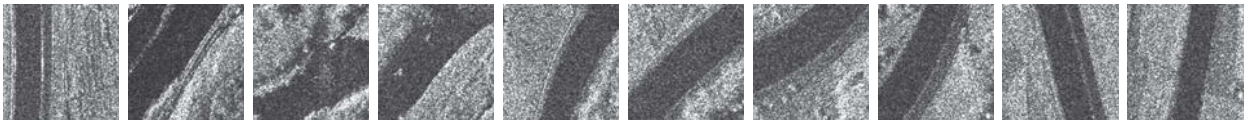
(c) **C07- Urban-03.** Venice (Italy).



(d) **C08- Urban-04.** Las-Vegas (USA).

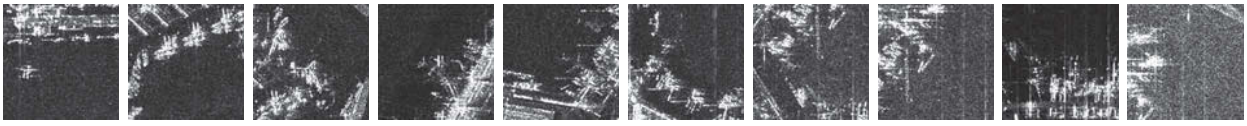


(e) **C09- Urban water-channel.** Venice (Italy).

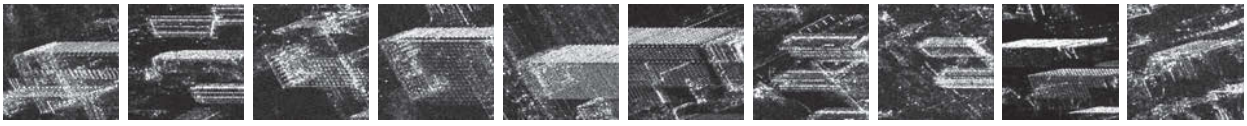


(f) **C10- Natural water-channel.** Maribor (Slovania).

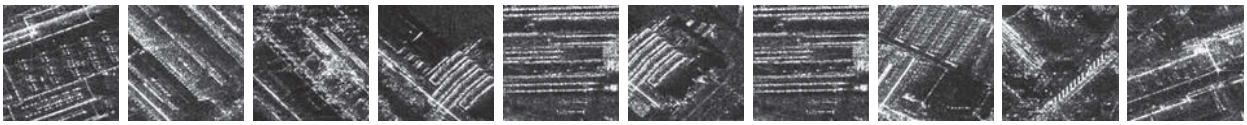
(g) **C11- Urban coastline.** Venice (Italy).



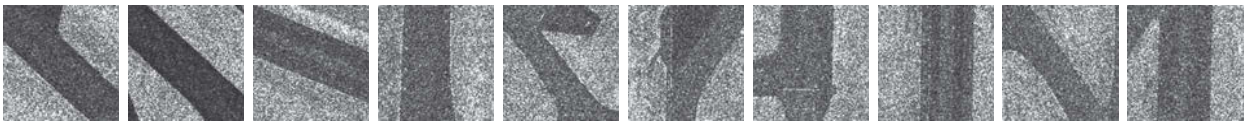
(h) **C12- Aircraft stands.** Hong Kong (P.R. China), Stockholm (Sweden).



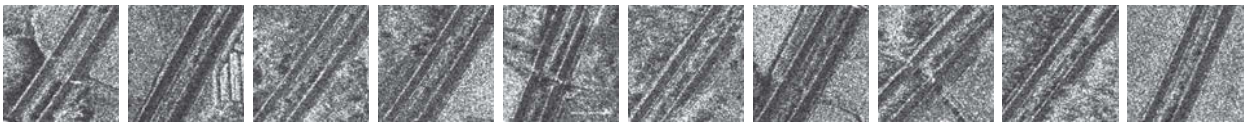
(i) **C13- Skyscrapers.** Tokyo (Japan), Vancouver (Canada), Bangkok (Thailand).



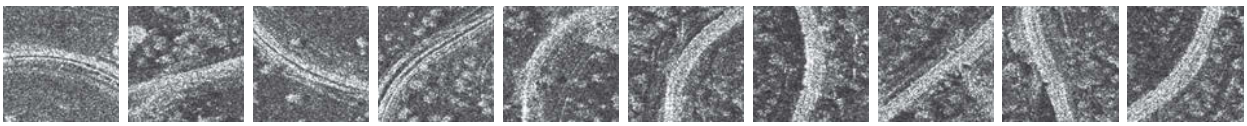
(a) **C14- Industrial complex.** Anshan (P.R. China).



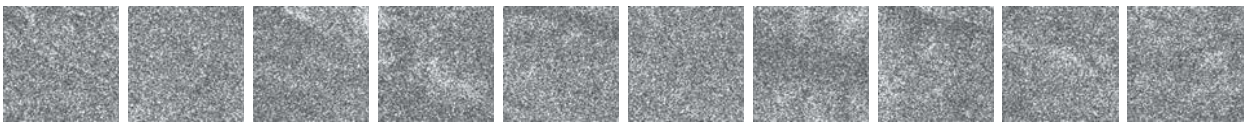
(b) **C15- Airstrip.** Hong Kong (P.R. China), Stockholm (Sweden), Gilching (Germany).



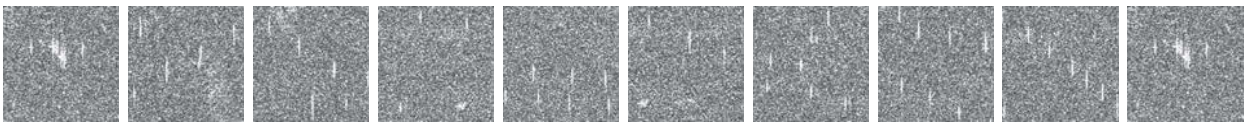
(c) **C16- Highway.** Stockholm (Sweden).



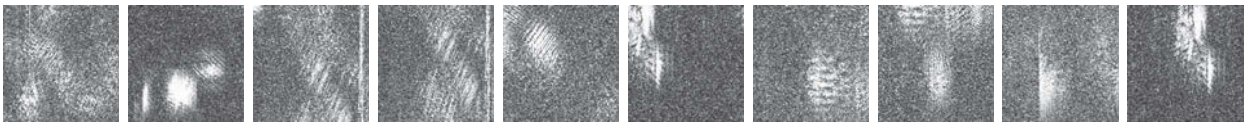
(d) **C17- Train tracks.** Tehachapi (USA).



(e) **C18- Sandy plains.** Giza Necropolis (Egypt).



(f) **C19- Buoy in Sea.** Venice (Italy).



(g) **C20- Ambiguities.** Venice (Italy), Hong Kong (P.R. China), Singapore.

Figure B.1: Examples of the SAR image patches (of the size 200×200 pixels) in different objects / texture categories. The total number of image-patches in each category is 100. The above shown images demonstrates the diversity of structures and objects in each category. It can be seen in C09, C10, the orientation of water channel is varying. Similar is the case with the C04, C15 and C17, where the texture is similar with objects oriented at different angles as well as exhibiting different backscattering behavior. The geographical locations from which various image-patches for each category are extracted is also mentioned.

Appendix C

Method of logarithmic-cumulants

In chapter 6, we introduced the use of logarithmic-cumulants as a statistical measure to generate a feature descriptor from the amplitude-envelope of the fractional Fourier transform coefficients as well as Gabor filtered images. This idea been motivated from the method of logarithmic-cumulants (MoLC) proposed in Nicolas [2002], as an alternative to the the maximum likelihood estimation (MLE) and method of moments (MoM) to solve a problem of parameter estimation for densities defined on $\mathbb{R}^+$. The MLE approach has already been introduced in section 4.2.1 in perspective to texture modal parameter estimation in Bayesian inference. In this appendix, we present a basic overview on MLE on the perspective of probability density function (pdf)parameter estimation, as well as the MoM. Later the MoLC is introduced.

The maximum likelihood estimation The MLE is a standard approach to the parameter estimation and inference in statistics. The MLE method chooses the parameter values that provides the highest value of the likelihood function by finding the root of the derivative of the likelihood function. Let us assume a set of random variables, $X = \{x_1, x_2, \cdots, x_k\}$, which has a pdf $p(X|\boldsymbol{\theta})$ with model parameter vector $\boldsymbol{\theta}$ with number of parameters $= p$. If all of the k individual observations, x_i are statistically independent of each other, the pdf of the given data X can be expressed as multiplication of the pdfs of the individual observations as follows (Myung [2003]):

$$p(X = (x_1, x_2, \cdots, x_k)|\boldsymbol{\theta}) = p_1(x_1|\theta)p_2(x_2 \mid \theta)\cdots p_k(x_k \mid \theta) = \prod_{i=1}^{k} p_x(x_i). \tag{C.1}$$

Now the likelihood function is defined by reversing the role of the observation vector X and the parameter vector θ in $p(X|\theta)$ as follows:

$$\mathcal{L}(\theta \mid x_1, x_2, \cdots, x_k) = p(X = (x_1, x_2, \cdots, x_k) \mid \theta) = \prod_{i=1}^{k} p_x(x_i). \tag{C.2}$$

Here $\mathcal{L}(\cdot)$ represents the likelihood of the parameter vector θ, given the observed data X. In MLE approach, one seeks to maximize this likelihood function. In practice, its more convenient to work with the logarithm of the likelihood function. Taking the natural logarithm of this likelihood function, the log-likelihood function can be equivalently written as:

$$\ln \mathcal{L}(\theta|x_1, x_2, \cdots, x_k) = \sum_{i=1}^{k} \ln p_x(x_i). \tag{C.3}$$

To maximize this log-likelihood function, partial derivative for the each element in θ parameter vector is taken, to obtain a system of p equations with p unknowns can be obtained, as follows:

$$\frac{\partial\left(\sum_{i=1}^{k}\ln p_x(x_i)\right)}{\partial\theta_j}=0,\quad j=1,2,\cdots,p. \tag{C.4}$$

The solution relies on the existence of the partial derivatives of the log-likelihood function. This approach attain the Cramer-Rao boundary, when feasible to apply (Nicolas [2002]). But it is difficult to obtain the partial derivative for complicated SAR-relevant distribution such as $\mathcal{K}$ distribution (Oliver [1993]).

The method of moments The MoM is a parameter estimation method which matches the theoretical moments with the sample moments. Let's assume X as a set of random variables with cumulative distribution function (cdf) $F(x)$ and the pdf $p(x)$. With $dF(x)/dx = p(x)$, its moments of order n (i.e. μ_n) are defined by its mathematical expectation $\mathbb{E}(X^n)$ as follows:

$$\mathbb{E}(X^n)=\int_{-\infty}^{+\infty}x^n dF(x)=\int_{-\infty}^{+\infty}x^n p(x)dx. \tag{C.5}$$

Now the sample moments of the observation in the random variable X are equated with the unobservable population moments of the distribution to solve for the unknown variables.

To illustrate the parameter estimation using the MoM, we consider the case of a simple and well-known two parameter Gamma distribution ($\mathcal{G}[\beta, L]$) widely used in characterizing speckle pdf in SAR images. Here the unknown parameters are: β, the scale parameter and L, the shape parameter. Its pdf is given as follows:

$$p(x|\beta,L)=\mathcal{G}\left[\beta,L\right](x)=\frac{1}{\Gamma(L)}\frac{L}{\beta}\left(\frac{Lx}{\beta}\right)^{L-1}\mathrm{e}^{-\frac{Lx}{\beta}},\quad \text{with } L>0. \tag{C.6}$$

Substituting the equation C.6 in the equation C.5 will provide the the first two population moments (analytical expression for the moments), which can be written as follows:

$$\mu_1=L\beta,\quad \mu_2=\beta^2L(L+1). \tag{C.7}$$

A system of equations is obtained by equating the sample moments (m_1 and m_2) and the analytical population moments (μ_1 and μ_2). Solving this system of equation will provide us the estimates for the unknown parameters (β and L) as follows:

$$\beta=\frac{m_2-m_1^2}{m_1},\quad L=\frac{m_1^2}{m_2-m_1^2}. \tag{C.8}$$

The use of MoM is justified when it is not possible to obtain the parameter estimates by the MLE approach. However, the variance of the estimates is higher in MoM as compared to the MLE.

The method of logarithmic-cumulants In the chapter 5, we presented the use of Fourier transform as an important signal processing tool for feature-descriptor generation. That includes the direct use of Fourier transform (section 5.3), the generalization of the Fourier transform (i.e. the fractional Fourier transform) (section 5.4), and in Gabor filter bank based approach (section 5.5) for computing the multiplication of signal by means of convolution in Fourier domain. However, the Fourier transform is just not a widely applicable signal processing tool, but it is also a fundamental notion in probability theory and statistical analysis.

In the classical statistical approach (Stuart & Ord [1987]), to define a random phenomenon by a pdf $p_x(u)$, the moment generating function or the first characteristic function $\Phi_x(v)$ is given by its Fourier transform $\mathcal{F}(\cdot)$, given as follows (Métivier [1972]):

$$\Phi_x(v) = \mathcal{F}\left[p(u)\right](v) = \int_{-\infty}^{\infty} \mathrm{e}^{jvu} p_x(u) du. \tag{C.9}$$

This moment generating function determines the moments when they exists. The moments μ_n of order n can be generated by differentiating this function to n^{th} order, written as follows:

$$\mu_n = (-j)^n \left.\frac{d^n \Phi_x(v)}{dv^n}\right|_{v=0} \tag{C.10}$$

The fundamental theorem of the theory of characteristics function, i.e. the inversion theorem, proves that a pdf $p_x(u)$ can be uniquely determined by its first characteristic function (Stuart & Ord [1987]). The cumulative function, also termed as the second characteristic function $\Psi_x(v)$ is defined as the logarithm of the first characteristic function:

$$\Psi_x(v) = \log(\Phi_x(v)). \tag{C.11}$$

It generates cumulants in the same way as the first characteristic function generates the moments. Thus the cumulants κ_n of order n are generated by differentiating the second characteristic function to n^{th} order, written as follows:

$$\kappa_n = (-j)^n \left.\frac{d^n \Psi_x(v)}{dv^n}\right|_{v=0} \tag{C.12}$$

We consider again the case of the Gamma distribution ($\mathcal{G}[\beta, L]$) for analyzing its distribution with the help of moment generating function. Here β is the scale parameter and L is the shape parameter. Same as equation C.6, its pdf is given as follows:

$$\mathcal{G}\left[\beta, L\right](u) = \frac{1}{\Gamma(L)} \frac{L}{\beta} \left(\frac{Lu}{\beta}\right)^{L-1} \mathrm{e}^{-\frac{Lu}{\beta}}, \quad \text{with } L > 0. \tag{C.13}$$

Now, the first characteristics function of Gamma distribution will be given by the Fourier transform of its pdf as follows:

$$\Phi(\nu) = \left(\frac{L}{\beta}\right)^L \frac{\mathrm{e}^{jL\arctan\left(\frac{\nu\beta}{L}\right)}}{\left(\nu^2 + \frac{L^2}{\beta^2}\right)^{\frac{L}{2}}} \tag{C.14}$$

In order to estimate the 2 parameters ν and L, we should obtain a system of equations formulated by the moments as a function of unknown variables ν and L by carrying out the differentiation of the equation C.14 to the first and second order. This should lead us to the equation C.7. Even in case of simple model such as Gamma distribution, this system of equation generates rather complicated expressions.

An alternative approach to obtain an easier solution to such problem was proposed in Nicolas [2002], by introducing the concept of *second-kind statistics*. That is why, we can refer to the statistical approach presented in earlier part of this section as the *first-kind statistics*. The approach *second-kind statistics*, can also be termed as the method of log-cumulants (MoLC), essentially proposes the expression for 'second kind characteristic functions', 'second kind moments (log-moments)' and 'second kind cumulants (log-cumulants)'. In the *second-kind statistics*, for the generation of the first characteristic function $\phi_x(v)$ (of second-kind), Mellin transform $\mathcal{M}(\cdot)$ of the pdf is used instead of the Fourier transform $\mathcal{F}(\cdot)$. This approach is applicable only to set of positive-valued random variable, X, with the pdf $p_x(u)$ defined for $u \in \mathbb{R}^+$. With the use of Mellin transform, the first characteristic function of second-kind, can be defined as:

$$\phi_x(v) = \mathcal{M}\left[p(u)\right](s) = \int_0^{+\infty} u^{s-1} p_x(u) du, \tag{C.15}$$

provided the integral in the above equation converges. Here, $s = a + ib$ denotes the complex variable, bounded to the strip l delimited by two parallel lines on the secondary axis where the Mellin transform is defined. For $s = 1$, the fundamental property $\phi_x(s)|_{s=1} = \int_0^{\infty} p_x(u) du = 1$ is satisfied (Li *et al.* [2011]). With these conditions, it is also possible to write the relations for the inverse Mellin transform, in order to determine the probability density function from the first characteristic function of second-kind as follows:

$$p_(u) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} u^{-s} \phi_x(s) ds, \quad c > l. \tag{C.16}$$

The equations C.15 and C.16 are also called the Mellin transform pair. Analogous to the equation C.10, the *second kind moments*, $\tilde{\mu}_n$ of order n can be defined by differentiating the first characteristic function of second-kind to n^{th} order as follows:

$$\tilde{\mu}_n = \left. \frac{d^n \phi_x(s)}{ds^n} \right|_{s=1} \tag{C.17}$$

Staying with the analogy in classical statistics, or the first kind statistics, the second characteristic function of second-kind can be defined as the logarithm of the first characteristic function of second-kind:

$$\psi_x(s) = \log(\phi_x(s)). \tag{C.18}$$

The n^{th} order derivatives of the second characteristic function of second-kind, evaluated at $s = 1$ will provide us the equations for the n^{th} order cumulants:

$$\tilde{\kappa}_n = \left. \frac{d^n \psi_x(s)}{ds^n} \right|_{s=1} \tag{C.19}$$

Now, again considering the case of the Gamma distribution for the random variables defined on $\mathbb{R}$, e.g. amplitude SAR images, the second characteristic function of second-kind for this distribution can be written, by the use of Mellin transform as follows (Oberhettinger [1974]):

$$\phi_x(s) = \mu^{s-1} \frac{\Gamma(L + s - 1)}{L^{s-1}\Gamma(L)} \tag{C.20}$$

The following system of equations for the cumulants of second kind can be derived from the equation C.20, by taking its first and second order derivative, with $\Psi(\nu, \cdot)$ as the ν^{th} order polygamma function as follows:

$$\tilde{\kappa}_1 = \Psi(0, L) + \ln(\mu) - \ln(L), \quad \tilde{\kappa}_2 = \Psi(1, L). \tag{C.21}$$

Here, $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ are the estimates of the first two cumulants of second-kind (or simply the logarithmic-cumulant, in this case, the logarithmic-mean and the logarithmic-variance). The first two logarithmic-cumulants for a given data, $I = \{r_1, r_2, \cdots, r_N\}$ can be estimated as follows:

$$\tilde{\kappa}_1 = \frac{1}{N}\sum_{k=1}^{N} \ln r_k, \quad \tilde{\kappa}_2 = \frac{1}{N-1}\sum_{k=1}^{N} (\ln r_k - \kappa_1)^2, \tag{C.22}$$

After obtaining these logarithmic-cumulants estimates, the system of equation given in equation C.21 is rather easier to solve than the case of the classical statistical approach of moments, to obtain the estimates for the unknown parameters μ and L for the Gamma distribution.

Based on the MoLC, it is possible to carry out the parameter estimation for many other complicated SAR-relevant pdfs mentioned in section 2.3.1. In some of the mentioned distribution, it is not possible to define the moments of certain order and also the maximum likelihood approach is very cumbersome to be applied to the practical applications. Thus the MoLC provides an efficient alternative to tackle the problem of parameter estimation for complicated distributions.

Nomenclature

ABM Autobinomial Modal

ANN Artifical Neural Networks

ASM Angular Second Moment

BW Bandwidth

CP Cluster Prominence

CS Cluster shade

DLR German Aerospace Center

EM Expectation Maximization

EPOS Edge Preserving Optimized Speckle-Filter

ESA European Space Agency

FM Frequency Modulation

FrFT The Fractional Fourier Transform

FT Fourier Transform

GEF Gabor Elementory Functions

GGD Generalized Gaussian Distribution

GLCM Gray Level Co-occurence Matrices

GMRF Gauss markov Random Fields

GRF Gibbs Random Fields

ICM Iterative Conditional Modes

IDM Inverse Difference Moment

IEM Iterative Evidence Maximization

JPEG Joint Photographic Experts Group

JTFA Joint Time Frequency Analysis

K-NN K-Nearest Neighbours

LFM Linear Frequency Modulation

MAP Maximum *a posteriori*

MCMC Markov Chain Monte Carlo

MLE Maximum Likelihood Estimatior

MMSE Minimum Mean Squared Error

MoLC Method of Logarithmic-Cumulants

MoM Method of Moments

MPEG Moving Pictues Experts Group

MSLD Multiple Sub-Look Decomposition

PA Producer's Accuracy

PRF Pulse Repetition Frequency

RAR Real Aperture Radar

SAR Synthetic Aperture Radar

SLC Single Look Complex

STFT Short Time Fourier Transform

SVM Support Vector Machines

TFA Time Frequency Analysis

UA User's Accuracy

VHR Very High Resolution

WVD Wigner-Ville Distribution

XML Extensible Markup Language

References

Ahmed, N. & Rao, K.R. (1975). *Orthogonal Transforms for Digital Signal Processing*. Springer-Verlag, Berlin ; New York :. 89

Albgretsen, F. (2008). Statistical Texture Measures Computed from Grey Level Co-occurrence Matrices. http://www.uio.no/studier/emner/matnat/ifi/INF4300/h08/undervisningsmateriale/glcm.pdf. 85, 86, 87

Almeida, L. (1994). The Fractional Fourier Transform and Time-Frequency Representations. *Signal Processing, IEEE Transactions on*, **42**, 3084 –3091. 93, 94

Arnaud, A. (1999). Ship Detection by SAR Interferometry. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, vol. 5, 2616 –2618 vol.5. 34

Bamler, R. & Eineder, M. (2008). The Pyramids of Gizeh Seen by TerraSAR-X - A Prime Example for Unexpected Scattering Mechanisms in SAR. *Geoscience and Remote Sensing Letters, IEEE*, **5**, 468–470. 8

Bamler, R. & Schaettler, B. (1993). *SAR Data Acquisition and Image Formation in: Geocoding:ERS-1 SAR Data and Systems,Ed. G. Schreier.*. Wichman, Karlsruhe. 12, 16

Baraldi, A. & Parmiggiani, F. (1995). An Investigation of the Textural Characteristics Associated with Gray Level Cooccurrence Matrix Statistical Parameters. *Geoscience and Remote Sensing, IEEE Transactions on*, **33**, 293 –304. 42, 84

Barbarossa, S. & Farina, A. (1992). Detection and Imaging of Moving Objects with Synthetic Aperture Radar. 2. Joint Time-Frequency Analysis by Wigner-Ville Distribution. *Radar and Signal Processing, IEE Proceedings F*, **139**, 89 –97. 41, 93

Barber, D.G. & Ledrew, E.F. (1991). SAR sea ice discrimination using texture statistics: a multivariate approach. *Photogrammetric Engineering and Remote Sensing*, **57**, 385–395. 86

Bayes, T. (1763). An Essay Towards Solving a Problem in the Doctrine of Chances. *Phil. Trans. of the Royal Soc. of London*, **53**, 370 – 418. 59

Bertrand, J. & Bertrand, P. (1996). The Concept of Hyperimage in Wide-Band Radar Imaging. *Geoscience and Remote Sensing, IEEE Transactions on*, **34**, 1144 –1150. 2, 34

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society*, **36**, 192–236. 43, 63

Besag, J. (1989). Towards Bayesian Image Analysis. *Journal of Applied Statistics*, **16**, 395–407. 61

Bovik, A., Clark, M. & Geisler, W. (1990). Multichannel Texture Analysis using Localized Spatial Filters. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **12**, 55 –73. 42, 103, 106

Box, G.E.P. & Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces)*. Wiley. 61

Bultheel, A., Martnez, H., Bultheel, A., Martnez, H., Bultheel, A. & Martnez, H. (2002). A Shattered Survey of the Fractional Fourier Transform. 94

Caelli, T. (1985). Three Processing Characteristics of Visual Texture Segmentation. *Spatial Vision*, **1**, 19 – 30. 101

CAF-DLR (2009). *TerraSAR-X Ground Segment: Basic Product Specification Document,* TX-GS-DD-3302. 36, 151

Capus, C. & Brown, K. (2003). Short-Time Fractional Fourier Methods for the Time-Frequency Representation of Chirp Signals. *Acoustical Society of America Journal*, **113**, 3253 – 3263. 94

Chellappa, R. & Jain, A. (1993). *Markov Random Fields: Theory and Application*. Academic Press. 63

Chen, V., Ling, H. & Miceli, W. (2003). Editorial: Time-frequency analysis for synthetic aperture radar and feature extraction. *Radar, Sonar and Navigation, IEE Proceedings* -, **150**, 201 – 202. 2

Clausi, D. & Yue, B. (2004). Comparing Cooccurrence Probabilities and Markov Random Fields for Texture Analysis of SAR Sea Ice Imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, **42**, 215 – 228. 84, 86

Clausi, D.A. (2000). Comparison and Fusion of Co-occurrence, Gabor and MRF Texture Features for Classification of SAR Sea-Ice Imagery. *Atmosphere-Ocean*, **39**, 183–194. 84

Cohen, L. (1989). Time-frequency distributions-a review. *Proceedings of the IEEE*, **77**, 941 –981. 139

Colin, E., Tria, M., Titin-Schnaider, C., Ovarlez, J. & Benidir, M. (2004). SAR Imaging using Multidimensional Continuous Wavelet Transform and Applications to Polarimetry and Interferometry. *International Journal of Imaging Systems and Technology*, **14**, 206–212. 2, 35

Cover, T. & Hart, P. (1967). Nearest Neighbor Pattern Classification. *Information Theory, IEEE Transactions on*, **13**, 21 –27. 115

Cross, G.R. & Jain, A.K. (1983). Markov Random Field Texture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-5**, 25 –39. 63, 70

Curlander, J.C. & McDonough, R.N. (1992). *Synthetic Aperture Radar: Systems and Signal Processing*. Wiley. 8, 12, 13

Datcu, M., Seidel, K. & Walessa, M. (1998). Spatial Information Retrieval from Remote-Sensing Images. I. Information Theoretical Perspective. *Geoscience and Remote Sensing, IEEE Transactions on*, **36**, 1431 –1445. 43, 65

Datcu, M., Stoichescu, D.A., Seidel, K. & Iorga, C. (2004). Model Fitting and Model Evidence for Multiscale Image Texture Analysis. In *24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 735, 35 – 42. 74

Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia, PA: SIAM. 76

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38. 65

Derin, H. & Elliott, H. (1987). Modeling and Segmentation of Noisy and Textured Images using Gibbs Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-9**, 39–55. 63, 64

Do, M. & Vetterli, M. (2002). Wavelet-based Texture Retrieval using Generalized Gaussian Density and Kullback-Leibler Distance. *Image Processing, IEEE Transactions on*, **11**, 146 –158. 43, 76, 77, 78, 131

Dunn, D., Higgins, W. & Wakeley, J. (1994). Texture Segmentation using 2-D Gabor Elementary Functions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **16**, 130 –149. 103

Eineder, M., Breit, H. & Fritz, T. (2008). Tips and Tricks for Spectral, Radiometric and Geometric Interpretation of TerraSAR-X Products. *3 TerraSAR-X Science Meeting*. 19, 20

Eineder, M., Adam, N., Bamler, R., Yague-Martinez, N. & Breit, H. (2009). Spaceborne Spotlight SAR Interferometry With TerraSAR-X. *Geoscience and Remote Sensing, IEEE Transactions on*, **47**, 1524 –1535. 20, 21, 36

Elachi, C. (1980). Spaceborne Imaging Radar: Geologic and Oceanographic Applications. *Science*, **209(4461)**, 1073 –1082. 30

Farrokhnia, F. & Jain, A. (1991). A Multi-Channel Filtering Approach to Texture Segmentation. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, 364 –370. 42, 104, 136

Fix, E. & Jr (1951). Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties. Tech. Rep. Project 21-49-004, Report Number 4, USAF School of Aviation Medicine, Randolf Field, Texas. 115

Franceschetti, G. & Lanari, R. (1999). *Synthetic Aperture Radar Processing*. CRC Press. 11

Frery, A., Muller, H.J., Yanasse, C. & Sant'Anna, S. (1997). A Model for Extremely Heterogeneous Clutter. *Geoscience and Remote Sensing, IEEE Transactions on*, **35**, 648 –659. 25

Frost, V.S., Stiles, J.A., Shanmugan, K.S. & Holtzman, J.C. (1982). A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **PAMI-4**, 157 –166. 45

Gabor, D. (1946). Theory of Communication. *J. Inst. Elect. Eng.*, **93**, 429–457. 103

Gaskill, J.D. (1978). *Linear Systems, Fourier Transforms, and Optics*. Wiley, New York :. 89

Geman, S. & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **PAMI-6**, 721 –741. 47

Goodman, J.W. (1976). Some Fundamental Properties of Speckle. *J. Opt. Soc. Am.*, **66**, 1145–1150. 22

Guarnieri, A. (2005). Adaptive Removal of Azimuth Ambiguities in SAR Images. *Geoscience and Remote Sensing, IEEE Transactions on*, **43**, 625 – 633. 18, 34, 95

Hagg, W. & Sties, M. (1994). Efficient Speckle Filtering of SAR Images. In *Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International*, vol. 4, 2140 –2142 vol.4. 45

Hammersley, J.M. & Clifford, P. (1971). Markov Field on Finite Graphs and Lattices. 63

Haralick, R. (1979). Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, **67**, 786 – 804. 42, 84

Haralick, R.M., Shanmugam, K. & Dinstein, I. (1973). Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on*, **SMC-3**, 610 –621. 42, 84, 87

Hassner, M. & Sklansky, J. (1980). The Use of Markov Random Fields as Models of Texture. *Computer Graphics and Image Processing*, **12**, 357 – 370. 43, 63

Hebar, M., Gleich, D. & Cucej, Z. (2009). Autobinomial Model for SAR Image Despeckling and Information Extraction. *Geoscience and Remote Sensing, IEEE Transactions on*, **47**, 2818 –2835. 1, 47, 65, 70

Hein, A. (2004). *Processing of SAR data*. Springer. 12, 16

Hinz, S., Meyer, F., Eineder, M. & Bamler, R. (2007). Traffic Monitoring with Spaceborne SAR-Theory, Simulations, and Experiments. *Comput. Vis. Image Underst.*, **106**, 231–244. 34

Hunt, M., Lennig, M. & Mermelstein, P. (1980). Experiments in Syllable-based Recognition of Continuous Speech. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '80.*, vol. 5, 880 – 883. 90

Ion, A.L. (2009). Methods for Knowledge Discovery in Images. *Information Technology and Control, Journal on*, **48**, 43 – 50. 87

Jain, A.K. & Vailaya, A. (1996). Image Retrieval using Color and Shape. *Pattern Recognition*, **29**, 1233–1244. 2, 41

Jaynes, E.T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press. 58

John G. Proakis, D.S., Dimitris G. Manolakis (2006). *Digital Signal Processing, Principles, Algorithms, and Applications*. Pearson Education. 89

JPEG-Committee (2013). Joint Photographic Experts Group (JPEG) homepage. `http://www.jpeg.org`. 42, 76

Julesz, B. (1981). Textons, the Elements of Texture Perception, and Their Interactions. *Nature*, **290**, 91–97. 83, 101, 106

Kamarainen, J.K., Kyrki, V. & Kalviainen, H. (2006). Invariance Properties of Gabor Filter-Based Features- Overview and Applications. *Image Processing, IEEE Transactions on*, **15**, 1088 –1099. 107

Kandaswamy, U., Adjeroh, D. & Lee, M. (2005). Efficient Texture Analysis of SAR Imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, **43**, 2075 – 2083. 43, 84, 104, 128

Kashyap, R. & Chellappa, R. (1983). Estimation and Choice of Neighbors in Spatial-Interaction Models of Images. *Information Theory, IEEE Transactions on*, **29**, 60 – 72. 66

Kim, A.J., III, J.W.F. & Willsky, A.S. (2003). Detection and Analysis of Anisotropic Scattering in SAR Data. *Multidimensional Systems and Signal Processing*, **14**, 49–82. 35

Kreyszig, E. (1979). *Statistische Methoden und ihre Anwendungen*. Vandenhoeck & Ruprecht, Goettingen. 61

Kuan, D., Sawchuk, A., Strand, T. & Chavel, P. (1987). Adaptive Restoration of Images with Speckle. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, **35**, 373 – 383. 46

Kuan, D.T., Sawchuk, A.A., Strand, T.C. & Chavel, P. (1985). Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **PAMI-7**, 165 –177. 44

Kuruoglu, E. & Zerubia, J. (2004). Modeling SAR Images with a Generalization of the Rayleigh Distribution. *Image Processing, IEEE Transactions on*, **13**, 527 – 533. 24

Lee, J.S. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **PAMI-2**, 165 –168. 44, 45

Lee, J.S. & Jurkevich, I. (1989). Segmentation of SAR Images. *Geoscience and Remote Sensing, IEEE Transactions on*, **27**, 674 – 680. 82

Li, F. & Johnson, W. (1983). Ambiguities in Spacebornene Synthetic Aperture Radar Systems. *Aerospace and Electronic Systems, IEEE Transactions on*, **AES-19**, 389 –397. 18

Li, H.C., Hong, W., Wu, Y.R. & Fan, P.Z. (2010). An Efficient and Flexible Statistical Model Based on Generalized Gamma Distribution for Amplitude SAR Images. *Geoscience and Remote Sensing, IEEE Transactions on*, **48**, 2711 –2722. 24, 131

Li, H.C., Hong, W., Wu, Y.R. & Fan, P.Z. (2011). On the Empirical-Statistical Modeling of SAR Images With Generalized Gamma Distribution. *Selected Topics in Signal Processing, IEEE Journal of*, **5**, 386 –397. 160

Li, S.Z. (1995). *Markov random field modeling in computer vision*. Springer-Verlag, London, UK, UK. 63, 65

Li, T. & Ogihara, M. (2006). Toward Intelligent Music Information Retrieval. *Multimedia, IEEE Transactions on*, **8**, 564 – 574. 90

Long, T., Liu, L. & Ding, Z. (2011). Strategy of Doppler Centroid Estimation in Synthetic Aperture Radar. *Radar, Sonar Navigation, IET*, **5**, 279 –287. 96, 97, 101

Lopes, A., Nezry, E., Touzi, R. & Laur, H. (1990a). Maximum A Posteriori Speckle Filtering And First Order Texture Models In Sar Images. In *Geoscience and Remote Sensing Symposium, 1990. IGARSS '90. 'Remote Sensing Science for the Nineties'., 10th Annual International*, 2409 –2412. 46, 47

Lopes, A., Touzi, R. & Nezry, E. (1990b). Adaptive Speckle Filters and Scene Heterogeneity. *Geoscience and Remote Sensing, IEEE Transactions on*, **28**, 992 –1000. 45

MacKay, D.J. (1991). Bayesian Interpolation. *Neural Computation*, **4**, 415–447. 61, 68

MacQueen, J.B. (1967). Some Methods for Classification and Analysis of MultiVariate Observations. In L.M.L. Cam & J. Neyman, eds., *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 281–297, University of California Press. 115

Madsen, S. (1989). Estimating the Doppler Centroid of SAR Data. *Aerospace and Electronic Systems, IEEE Transactions on*, **25**, 134 –140. 36, 96, 97, 101

Mahalanobis, P.C. (1936). On the Generalised Distance in Statistics. In *Proceedings National Institute of Science, India*, vol. 2, 49–55. 115

Malik, J. & Perona, P. (1990). Preattentive Texture Discrimination with Early Vision Mechanisms. *Journal of the Optical Society of America A*, **7**, 923–932. 101

Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **11**, 674 –693. 76

Manjunath, B. & Ma, W. (1996). Texture Features for Browsing and Retrieval of Image Data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **18**, 837 –842. 42, 104, 106, 136

Massonnet, D. & Souyris, J.C. (2008). *Imaging with Synthetic Aperture Radar*. EFPL Press. 41, 95

Mcbride, A.C. & Kerr, F.H. (1987). On Namias's Fractional Fourier Transforms. *IMA Journal of Applied Mathematics*, **39**, 159–175. 93

Métivier, M. (1972). *Notions fondamentales de la théorie des probabilités.*. Dunod. 159

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**, 1087–1092. 67

Mitchell, T.M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1st edn. 116

Mittermayer, J. & Runge, H. (2003). Conceptual Studies for Exploiting the TerraSAR-X Dual Receive Antenna. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol. 3, 2140 – 2142. 36

Molina, D.E. (2011). *Advanced Methods for High-Resolution SAR Information Extraction: Data and User-Driven Evaluation Approaches for Image Information Mining*. Ph.D. thesis, Telecom ParisTech. 65

Moser, G., Zerubia, J. & Serpico, S. (2006). SAR Amplitude Probability Density Function Estimation based on a Generalized Gaussian Model. *Image Processing, IEEE Transactions on*, **15**, 1429 –1442. 24, 82

Myung, I.J. (2003). Tutorial on Maximum Likelihood Estimation. *J. Math. Psychol.*, **47**, 90–100. 157

Namias, V. (1980). The Fractional Order Fourier Transform and its Application to Quantum Mechanics. *IMA Journal of Applied Mathematics*, **25**, 241–265. 2, 93

Nicolas, J.M. (2002). Introduction aux statistiques de deuxième espèce: applications des logs-moments et des logscumulants à l'analyse des lois d'images radar. *Traitement du Signal*, **19**, 139 – 167. 2, 157, 158, 160

Novak, L.M., Owirka, G.J., Brower, W.S., & Weaver, A.L. (1997). The Automatic Target-Recognition System in SAIP. *The Lincon Laboratory Journal*, **10**, 187 – 202. 50

Oberhettinger, F. (1974). *Tables of Mellin Transforms.*. New York: Springer-Verlag. 160

Oliver, C. (1984). A Model for Non-Rayleigh Scattering Statistics. *Optica Acta: International Journal of Optics*, **31**, 701–722. 25

Oliver, C.J. (1993). Optimum Texture Estimators for SAR Clutter. *Journal of Physics D: Applied Physics*, **26**, 1824. 24, 158

Oliver, C.J. & Quegan, S. (1998). *Understanding Synthetic Aperture Images*. Artech House, Norwood, MA. 22, 23, 24, 43, 45, 62

Ozaktas, H., Arikan, O., Kutay, M. & Bozdagt, G. (1996). Digital Computation of the Fractional Fourier Transform. *Signal Processing, IEEE Transactions on*, **44**, 2141 –2150. 2, 94

Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York. 22, 59, 61, 131

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.*, 1065 –1076. 82

Picinbono, B. & Bouvet, M. (1984). Complex White Noises and Autoregressive Signals. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, vol. 9, 596 – 599. 73

Popescu, A., Gavat, I. & Datcu, M. (2008). SAR Image Time-Frequency Information Extraction based on Renyi Entropy. In *Proceedings of IWCIT-2008*. 41

Popescu, A., Patrascu, C., Singh, J. & Datcu, M. (2010). A Spectral Space-Variant Approach for Structure Indexing in Spotlight TerraSAR-X Data. In *Communications (COMM), 2010 8th International Conference on*, 169 –172. 41, 90

Popescu, A., Gavat, I. & Datcu, M. (2012). Contextual Descriptors for Scene Classes in Very High Resolution SAR Images. *Geoscience and Remote Sensing Letters, IEEE*, **9**, 80 –84. 3, 90, 124

Porcello, L.J., Massey, N.G., Innes, R.B. & Marks, J.M. (1976). Speckle Reduction in Synthetic-Aperture Radars. *J. Opt. Soc. Am.*, **66**, 1305–1311. 44

Qian, S. & Chen, D. (1993). Discrete Gabor Transform. *Signal Processing, IEEE Transactions on*, **41**, 2429 –2438. 103

Qian, S. & Chen, D. (1999). Joint Time-Frequency Analysis. *Signal Processing Magazine, IEEE*, **16**, 52 –67. 41, 93

Rees, W.G. & Satchell, M.J.F. (1997). The effect of median filtering on synthetic aperture radar images. *International Journal of Remote Sensing*, **18**, 2887–2893. 44

Riaz, F., Silva, F., Ribeiro, M. & Coimbra, M. (2012). Invariant Gabor Texture Descriptors for Classification of Gastroenterology Images. *Biomedical Engineering, IEEE Transactions on*, **PP**, 1. 104, 107

Ross, T.D., Worrell, S.W., Velten, V.J., Mossing, J.C. & Bryant, M.L. (1998). Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set. *Proc. SPIE*, **3370**, 566–573. 51

Ruanaidh, J.J.K.O. & Fitzgerald, W.J. (1996). *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag. 74

Rudin, W. (1976). *Principles of Mathematical Analysis*. 2nd ed. New York, McGraw-Hill. 23

Scheirer, E. & Slaney, M. (1997). Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 2, 1331 –1334 vol.2. 90

Schistad Solberg, A. & Jain, A. (1997). Texture Fusion and Feature Selection Applied to SAR Imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, **35**, 475 –479. 67

Schneider, R., Papathanassiou, K., Hajnsek, I. & Moreira, A. (2006). Polarimetric and Interferometric Characterization of Coherent Scatterers in Urban Areas. *Geoscience and Remote Sensing, IEEE Transactions on*, **44**, 971 – 984. 34

Schroder, M., Rehrauer, H., Seidel, K. & Datcu, M. (1998). Spatial Information Retrieval from Remote-Sensing Images. II. Gibbs-Markov Random Fields. *Geoscience and Remote Sensing, IEEE Transactions on*, **36**, 1446 –1455. 70

Shi, Z. & Fung, K. (1994). A Comparison of Digital Speckle Filters. In *Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International*, vol. 4, 2129 –2133 vol.4. 45

Shyu, C.R., Klaric, M., Scott, G., Barb, A., Davis, C. & Palaniappan, K. (2007). GeoIRIS: Geospatial Information Retrieval and Indexing System-Content Mining, Semantics Modeling, and Complex Queries. *Geoscience and Remote Sensing, IEEE Transactions on*, **45**, 839 –852. 1, 51

Sikora, T. (2001). The MPEG-7 Visual Standard for Content Description-an Overview. *Circuits and Systems for Video Technology, IEEE Transactions on*, **11**, 696 –702. 107, 136

Singh, J. & Datcu, M. (2010). Target Analysis using a 4-D Representation based on Sub-Aperture Decomposition of SAR Images. *Synthetic Aperture Radar (EUSAR), 2010 8th European Conference on*, 1 –3. 35, 36

Singh, J. & Datcu, M. (2012a). Automated Interpretation of Very-High Resolution SAR Images. In *Geoscience and Remote Sensing Symposium,2012 IEEE International, IGARSS 2012*. 3, 52, 53, 143

Singh, J. & Datcu, M. (2012b). Mining Very High Resolution Complex-valued SAR images using The Fractional Fourier Transform. *Synthetic Aperture Radar (EUSAR), 2012 9th European Conference on*, 1 –4. 42, 93, 101, 144

Singh, J. & Datcu, M. (2012c). SAR Target Analysis based on Multiple-Sublook Decomposition: A Visual Exploration Approach. *Geoscience and Remote Sensing Letters, IEEE*, **9**, 247 –251. 2, 34, 104, 144

Singh, J. & Datcu, M. (2012d). Use of Second-Kind Statistics for VHR SAR Image Retrieval. *Communications (COMM), 2012 9th International Conference on*, 1 –4. 42, 124, 131

Singh, J. & Datcu, M. (2013). SAR Image Categorization With Log Cumulants of the Fractional Fourier Transform Coefficients. *Geoscience and Remote Sensing, IEEE Transactions on*, **51**, 5273–5282. 2, 42, 93, 144

Singh, J., Soccorsi, M. & Datcu, M. (2009). Parametric versus Non-Parametric Complex Image Analysis. In *Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009*, vol. 3, III–9 –III–12. 41, 47, 65, 90, 115

Singh, J., Cui, S., Datcu, M. & Gleich, D. (2012). A Survey of Density Esimation for SAR Images. *The 2012 European Signal Processing Conference (EUSIPCO-2012)*. 25

Sivia, D.S. (1996). *Data Analysis: A Bayesian Tutorial*. Oxford, U.K.: Clarendon. 68

Smeulders, A., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000). Content-based Image Retrieval at the End of the Early Years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **22**, 1349 –1380. 1, 51

Soccorsi, M. (2010). *Parameter Estimation and Modeling of High-Resolution Synthetic Radara Data*. Ph.D. thesis, Telecom ParisTech. 47, 65, 73

Soh, L.K. & Tsatsoulis, C. (1999). Texture Analysis of SAR Sea Ice Imagery using Gray Level Co-occurrence Matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, **37**, 780 –795. 42, 84

Song, K.S. (2006). A Globally Convergent and Consistent Method for Estimating the Shape Parameter of a Generalized Gaussian Distribution. *Information Theory, IEEE Transactions on*, **52**, 510 –527. 77

Souyris, J.C., Henry, C. & Adragna, F. (2003). On the Use of Complex SAR Image Spectral Analysis for Target Detection: Assessment of Polarimetry. *Geoscience and Remote Sensing, IEEE Transactions on*, **41**, 2725 – 2734. 2, 35

Spigai, M., Tison, C. & Souyris, J.C. (2011). Time-Frequency Analysis in High-Resolution SAR Imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, **49**, 2699 –2711. 2, 35, 36, 39, 98, 104

Stehman, S.V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, **62**, 77 – 89. 116

Story, M. & Congalton, R.G. (1986). Accuracy assessment - a user's perspective. *Photogrammetric Engineering and Remote Sensing*, **52**, 397 – 399. 116

Stuart, A. & Ord, J.K. (1987). *Kendall's advanced theory of statistics*. Griffin, London, 5th edn. 132, 159

Sun, H.B., Liu, G.S., Gu, H. & Su, W.M. (2002). Application of the Fractional Fourier Transform to Moving Target Detection in Airborne SAR. *Aerospace and Electronic Systems, IEEE Transactions on*, **38**, 1416 – 1424. 2, 42, 93, 94

Super, B. & Bovik, A. (1991). Three-dimensional orientation from texture using gabor wavelets. **1606**, 574–586. 103

Swain, M. & Ballard, D. (1990). Indexing via Color Histograms. In *Computer Vision, 1990. Proceedings, Third International Conference on*, 390 –393. 82

Szajnowski, W. (1977). Estimators of Log-Normal Distribution Parameters. *Aerospace and Electronic Systems, IEEE Transactions on*, **AES-13**, 533 –536. 24

Tison, C. & Fjortoft, R. (2008). Singular Point Characterization and Monitoring using Advanced TerraSAR-X Modes. *TerraSAR-X Science Team meeting, DLR, Oberpfaffenhofen*, 1 –12. 35, 36, 39

Tison, C., Nicolas, J.M., Tupin, F. & Maitre, H. (2004). A New Statistical Model for Markovian Classification of Urban Areas in High-Resolution SAR Images. *Geoscience and Remote Sensing, IEEE Transactions on*, **42**, 2046 – 2057. 23, 25, 132

Touzi, R. (2002). A Review of Speckle Filtering in the Context of Estimation Theory. *IEEE Transactions on Geoscience and Remote Sensing*, **40(11)**, 2392–2404. 1, 22, 25, 43, 62

Tuceryan, M. & Jain, A.K. (1998). Texture analysis. In C.H. Chen, L.F. Pau & P.S.P. Wang, eds., *Handbook of Pattern Recognition & Computer Vision*, 2, 207–248, World Scientific Publishing Co., Inc., River Edge, NJ, USA. 2, 41, 42, 43

Tupin, F. (2011). How Advanced Image Processing Helps For SAR Image Restoration and Analysis. *IEEE Geoscience and Remote Sensing Society Newsletter*, 10 – 17. 132

Tzanetakis, G. & Cook, P. (2002). Musical Genre Classification of Audio Signals. *Speech and Audio Processing, IEEE Transactions on*, **10**, 293 – 302. 90

Ulaby, F., Kouyate, F., Brisco, B. & Williams, T. (1986). Textural Information in SAR Images. *Geoscience and Remote Sensing, IEEE Transactions on*, **GE-24**, 235 –245. 30

Van de Wouwer, G., Scheunders, P. & Van Dyck, D. (1999). Statistical Texture Characterization from Discrete Wavelet Representations. *Image Processing, IEEE Transactions on*, **8**, 592 –598. 76

Varanasi, M.K. & Aazhang, B. (1989). Parametric Generalized Gaussian Density Estimation. *Acoustical Society of America Journal*, **86**, 1404 – 1415. 77

von der Linden, W., Preuss, R. & Dose, V. (1999). The prior-predictive value: A paradigm of nasty multidimensional integrals. *Maximum Entropy and Bayesian Methods, Eds. Norwell, MA: Kluwer*, 319–326. 68

Walessa, M. & Datcu, M. (2000). Model-based Despeckling and Information Extraction from SAR Images. *Geoscience and Remote Sensing, IEEE Transactions on*, **38**, 2258 –2269. 1, 47, 65, 67, 68